Vol. 53 FEBRUARY 1962 No. 1

Journal of Educational Psychology

CONTENTS

Intelligence at Adulthood: A Twenty-Five Year Follow-Up 1
KATHERINE P. BRADWAY AND CLARE W. THOMPSON

Foreign Language Learning Ability 15
PAUL PIMSLEUR, ROBERT P. STOCKWELL, AND ANDREW L. COMREY

Training Adults to Judge Children's Social Acceptance 27
HELEN R. MARSHALL

The Effects of Discrimination Pretraining on Textual Behavior 32
CAROLYN K. STAATS, ARTHUR W. STAATS, AND RICHARD E. SCHUTZ

Some Characteristics Related to Social Isolation of Second Grade Children 38
HARRIET H. COMMOSS

Personal Preference for Method as a Factor in Learning 43
NEWTON E. JAMES

Values of Medical School Students at Different Levels of Training 48
LEONARD V. GORDON AND IVAN N. MENSH

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

RAYMOND G. KUHLEN, Editor
Syracuse University

Advisory Editors

JOHN B. CARROLL
Harvard University
LEE J. CRONBACH
University of Illinois
FRANCIS J. DIVESTA
Syracuse University
CHESTER W. HARRIS
University of Wisconsin
ERNEST R. HILGARD
Stanford University
NICHOLAS HOBBS
George Peabody College for Teachers
DOROTHEA MCCARTHY
Fordham University
CARSON MCGUIRE
University of Texas
PAUL H. MUSSEN
University of California, Berkeley
T. ERNEST NEWLAND
University of Illinois
H. H. REMMERS
Purdue University
JULIAN C. STANLEY
University of Wisconsin
J. B. STROUD
State University of Iowa
DONALD E. SUPER
Teachers College, Columbia University
GEORGE G. THOMPSON
Ohio State University
BENTON J. UNDERWOOD
Northwestern University

The *Journal of Educational Psychology* publishes original investigations of problems of learning, teaching, and the psychological development of the individual. It contains articles on the following subjects: the psychology of school subjects; experimental studies of learning the development of interests, attitudes, personality, social relations, and vocational orientation, particularly (but not exclusively) as related to school adjustment or the impact of the school program upon development and adjustment; emotion, motivation, and character; mental development; and methods. The last includes tests, statistical techniques, and research techniques in cross-sectional and developmental studies. Studies published deal with all levels of education—preschool, elementary, secondary, college, and adult education—and with all age groups.

Manuscripts and correspondence on editorial matters should be addressed to the Editor: Raymond G. Kuhlen, Department of Psychology, Syracuse University, 123 College Place, Syracuse 10, New York.

All manuscripts should be submitted *in duplicate* to facilitate referral to advisory editors and must be prepared to conform to the style and procedures described in the *Publication Manual of the American Psychological Association.*

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts.* Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR, Acting Production Manager

Communications including subscriptions, orders of back issues, and changes of address should be addressed to the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D. C. Address changes must reach the APA by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $10.00 (Foreign $10.50). Single copies, $2.00.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

P. O. Box 1064, Richmond 8, Va.
and 1333 Sixteenth Street, N.W., Washington 6, D. C.
Second-Class postage paid at Richmond, Va.
Copyright, 1962 by the American Psychological Association, Inc.

JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 53 February 1962 Number 1

INTELLIGENCE AT ADULTHOOD: A TWENTY-FIVE YEAR FOLLOW-UP[1]

KATHERINE P. BRADWAY
Stanford University

AND

CLARE W. THOMPSON[2]
Washington State University

This paper reports findings from 111 Ss tested with the Stanford-Binet (S-B) at preschool and adolescence who were administered the S-B and WAIS at adulthood. Correlations of preschool IQs with adult S-B and full WAIS IQs are .59 and .64; of adolescent IQs with adult IQs, .85 and .80. Mean S-B IQ increase from adolescence to adulthood is 11 points, indicating that mental growth continues beyond 16 years. Males show more IQ gain after adolescence than do females ($p < .01$). Girls with high IQs increase least. Analysis of increases in percent passing S-B items shows more growth after adolescence in abstract reasoning and vocabulary than in rote memory and practical reasoning. Preschool verbal and memory items are better predictors than nonverbal items of both verbal and performance adult IQs. Pattern of individual differences in relative amounts of these abilities shows some stability over 25 years.

Attempts to understand human intelligence were based initially on studies of children. In contrast, this paper deals with aspects of intelligence which cannot be studied adequately from data limited to childhood intelligence. Specifically, our approach has been to study as young adults a group of individuals previously tested during the preschool years and during adolescence (Bradway 1944, 1945a, 1945b; Bradway & Robinson 1961; Bradway, Thompson, & Cravens 1958).

Binet and Simon developed the original intelligence test to objectify the criteria which educators had been using for many years in assessing the capacities of children for purposes of school placement. The several early American adaptations of this test culminated in the 1916 Stanford-Binet, which provided indices of adult intelligence. The standardization group, however, consisted largely of children; adult indices were based on extrapolation.

The urgency of the needs for screening and assignment created by World War I led to the creation of truly adult intelligence tests but, consistent with these needs, these tests were for group rather than individual administration. Later revisions of individual intelligence tests (e.g., the 1937 and 1960 Stanford and the 1939 Kuhlmann) extended the upper limit by providing more difficult test items. Moreover, somewhat older subjects were included in the standardization groups, although the upper age levels of these groups did not go above the adolescent years into adulthood.

No adequate instrument existed by which the clinician could measure the

[1] This investigation was supported by USPHS Research Grant M-1273 through the National Institute of Mental Health, National Institutes of Health.

[2] At Stanford University when participating in this research.

intelligence of the individual adult until the publication of the Wechsler-Bellevue in 1939. Wechsler's large standardization sample started with children of 12 years and included older age groups up to 65 years. His emphasis on adult intelligence resulted in departure from the use of mental age, a concept which had contributed immeasurably to the interpretation of childhood intelligence, but which introduced confusion in the assessment of adult intelligence. His term "IQ" is not a continuation of Stern's quotient of mental age divided by chronological age; it is more comparable to a standard score. Unlike the revisions of the Binet, all of which had been based on the concept that intellectual abilities reach their peak at from 14 to 16 years of age, indices of intelligence for the first time took account of the continuation of intellectual growth into the adult years. This view of intellectual growth has been increasingly confirmed by more recent studies (Bayley, 1955; Bayley & Oden, 1955; Miner, 1957; Owens, 1953) including the present one. Although Wechsler calls his concept of intelligence global, the structure of the Wechsler-Bellevue (and its later revision, the WAIS) is such that several kinds of tasks are separately measured and separate indices of verbal and nonverbal abilities are provided.

The clinician trying to answer questions about the intelligence of an adult needs information about the unique growth rates of different kinds of abilities. No cross-sectional sampling can provide such information accurately. On the other hand, longitudinal studies pose problems of practice effects and of the attrition which inevitably occurs over a period of time. The clinician is also confronted with the problem of how most meaningfully to express the measured intelligence of any given adult. Should it be an index which compares him with his own age group? One which locates him on his own course of development and decline? Or several which represent an assessment of his various kinds of mental processes? Moreover, each of these requires that procedures for equating means and variability be worked out. Answers to many of these questions are necessary before valid predictions about adult intelligence can be made.

PROCEDURE AND SUBJECTS

Information for some of these answers is provided by the longitudinal study we have recently completed (Bradway & Robinson, 1961; Bradway et al., 1958). A group of 111 adults aged 27 to 32 who had been tested 25 years earlier as part of the 1937 Stanford-Binet standardization sample of preschool children and again during their adolescent years (Bradway 1944, 1945a, 1945b) was administered the Stanford-Binet and WAIS in 1956.[3] The repetition of the Stanford-Binet permits analysis of improvement in the same tasks from adolescence to adulthood; the WAIS data permit comparison of child and adult measures of intelligence of the same individuals. Treatment of the data has included correlations between total scores, mental growth curves, an analysis of items for increase in percent passing from adolescence to adulthood, and an analysis of the predictive value of preschool items. This latter item analysis was based on: correlations between Stanford-Binet (Forms L and M) preschool items and both adolescent Stanford-Binet IQ and adult WAIS verbal and performance quotients, and comparisons between scores on special scales constructed from the several kinds of preschool items and later scores obtained on comparable adult WAIS scales. Because for one subject the WAIS was incomplete and educational attainment was not ascertained, and for another the 1931 Stanford-Binet IQ was invalid, the number of

[3] The authors are happy to express their gratitude to Maud A. Merrill and the late Lewis M. Terman for generously having made their standardization data available for follow-up; to Quinn McNemar, consultant and advisor throughout the study; and to Nancy M. Robinson who participated actively in the later stages of the study.

total subjects on which any one analysis is based varies from 109 to 111 depending upon which variables are involved in the analysis. No attempt was made in the two follow-ups to continue the procedure used in the initial examination of testing children within one month of a birthday; a larger deviation in the age distribution was a consequence.

RESULTS AND DISCUSSION

Constancy of the IQ

A finding of primary interest is that an individual's measured intelligence in relation to others of his same age is demonstrated to be more stable than previous studies suggest. These results are presented in Table 1. It will be noted that the correlation between preschool (1931) intelligence, and adult (1956) intelligence as measured by the full scale WAIS is .64. The measure of preschool intelligence was a combination of the two forms of the Stanford-Binet which became the 1937 Revision, and the measure of adult intelligence was a combination of the verbal and performance scales of the WAIS; thus the indices of intelligence at both ages were based on the most adequate measurements available when the data were collected. The reliability of each measurement was sufficiently high so that correcting for unreliability increases the correlation only slightly: from .64 to .69. The degree of correspondence over a 25-year period is not significantly different from the correlation of .65 (uncorrected) found between preschool and adolescent Stanford-Binet IQs after an interval of 10 years.

Also of note from Table 1 is the correlation of .80 between adolescent (1941) Stanford-Binet and adult (1956) WAIS. Although this is somewhat lower than the correlation of .85 between the adolescent and adult administrations of the Stanford-Binet, neither correlation is significantly different from the correlation of .83 between Stanford-Binet and WAIS administered to our subjects on the same day when they were adults. Apparently whatever the Stanford-Binet measures at preschool and at adolescence is still being measured by the WAIS at adulthood.

A comparison of the columns of Table 1 shows a close similarity, both in absolute amount and in trend, between the 1956 verbal WAIS and the 1956 full scale WAIS when each is correlated with the Stanford-Binet administered at the three age levels. The performance

TABLE 1

CORRELATIONS BETWEEN TEST SCORES AT THE SEVERAL ADMINISTRATIONS

Test	1941 S-B (L)	1956 S-B (L)	1956 WAIS Full	1956 WAIS Verbal	1956 WAIS Performance
1931 S-B (L and M)	.65	.59	.64	.60	.54
1941 S-B (L)		.85	.80	.81	.51
1956 S-B (L)			.83	.89	.46
1956 WAIS Full				.87	.84
1956 WAIS Verbal					.59

WAIS correlations, on the other hand, are both lower in value and opposite in trend. The reason for this reversal in trend needs further study to determine whether the results reflect an increase in specialization of intelligence with age or are due to an increase in verbal content of the Stanford-Binet at the higher levels.

Growth and Decline of Intellectual Abilities

Our data contribute to the accumulating evidence that intellectual potential does not cease to increase at 16 years in contradiction to the conclusions of Terman and others. Table 2 summarizes our data for comparison of obtained IQs. In view of the conformance between the 1931 and 1941 Stanford-Binet IQs and between the standard deviations at the three ages, the fact that the 1956 Stanford-Binet IQ exceeds the 1941 Stanford-Binet IQ by a statistically significant amount is interpreted as a consequence of intellectual growth in these subjects beyond 16 years. The increase of 11 points in IQ corresponds to a gain in mental age score of 20 months beyond expectancy. The finding of a continuation of mental growth into adulthood is consistent with the findings of Bayley (1955, 1957) and others (Bayley & Oden, 1955; Miner, 1957; Owens, 1953). The implications of the difference between present and previously held conclusions regarding mental growth into adulthood deserve consideration. Bayley (1955, p. 817) suggests alternative explanations including the new generations' having more opportunity to develop their intellectual capacities than did their parents. Inhelder and Piaget (1958, p. 337) have pointed out the probability that Greek children were behind our own in growth of logical thinking and suggest that the age at which formal thinking begins "may be a product of progressive acceleration of individual development under the influence of education."

TABLE 2

MEANS AND STANDARD DEVIATIONS OF THE TWO TESTS AT THE THREE ADMINISTRATIONS

Test	1931		1941		1956	
	M	*SD*	*M*	*SD*	*M*	*SD*
Chronological age	4.0	1.1	13.6	1.4	29.5	1.5
S-B IQ	112.8[a]	15.9[a]	112.3[b]	16.4[b]	123.6[b]	15.0[b]
WAIS IQ						
Full					108.9	11.0
Verbal					110.4	13.3
Performance					105.0	11.5

[a] Composite Form L and Form M.
[b] Form L.

The method of computing WAIS IQs assumes continuation of intellectual growth up to 34 years. The mean WAIS IQ of our group is lower in numerical value than expectancy based on the mean 1941 Stanford-Binet IQ. It will be noted, however, that the WAIS mean IQ is approximately one *SD* above 100, whereas the 1941 Stanford-Binet mean IQ is less than one *SD* above 100; so that in terms of relative deviations, the WAIS mean IQ is higher than the Stanford-Binet IQ, although not to a statistically significant degree. The verbal half of the WAIS approaches the 1941 Stanford-Binet results more closely than does the full scale, and the performance half less so.

The addition of the present data to those previously collected for this group of subjects gives us a third segment of the mental growth curve. Thus we can plot mental age against chronological age for the years 2 through 5, the years 11 through 15, and now the years 27 through 31. Two subjects were younger

TABLE 3

STANFORD-BINET SCORES FOR ADULTS BY YEAR OF AGE

1956 Chronological age	*N*	S-B Mean 1956 IQ	S-B Mean 1956 MA
27.0-27.9	16	121.3	18.2
28.0-28.9	25	125.5	18.9
29.0-29.9	23	126.8	19.0
30.0-30.9	20	123.5	18.5
31.0-31.9	24	121.0	18.1

and one was older than these years and are omitted from this section. Table 3 summarizes the current data for the Stanford-Binet.

Analysis of variance was applied to these data. The resulting *F* ratio was .594 for 4 and 103 degrees of freedom, whereas 2.45 would be required for the .05 level of significance. Of course, should the drop from age 29 to age 31 continue in direction and amount, the critical point of this mental growth curve would have been established, but there is no way to predict this from the present data.

Before we interpret what this lack of change during this 5-year period may mean, we must consider the possible ceiling effect. The 1937 Stanford-Binet has insufficient top for adults and therefore this flattening of the curve might be an artifact. To examine this possibility, a similar analysis was made of the results on the WAIS for these subjects, inasmuch as a low ceiling is not a limitation of that test. One of the subjects had a partial WAIS only and is omitted for this reason, reducing the degrees of freedom to 4 and 102, but not changing the value of 2.45 for the .05 level of significance against which the results are compared. Table 4 summarizes the data for the WAIS. Analysis of variance for these three scores again shows the variation with chronological age not to approach significance. The *F* ratios are, respectively, 1.40, 1.01, and 1.62.

Further study to determine the role of ceiling effect was made by analyzing only those cases showing no or few passes at the Superior Adult III level of the Stanford-Binet. Also, because our group of subjects is somewhat above average, a separate study was made of those cases clustering around IQ 100. These additional studies, however, served only to substantiate the results just reported. The point at which mental growth as measured by these instruments terminates and at which the curve assumes a downward direction is not established by these data.

Sex Differences

Tests such as the Stanford-Binet are intentionally constructed to eliminate or to minimize sex differences. Consequently most investigators, including the present ones in earlier studies, have given little attention to this variable. It is in this area, however, that some of the most cogent findings of the present study emerge. As shown in Table 5, on all three administrations of the Stanford-Binet, female subjects achieved a higher mean score than did males. On

TABLE 4

WAIS SCORES FOR ADULTS BY YEAR OF AGE

1956 Chronological age	*N*	Mean 1956 WAIS standard scores		
		Full	Verbal	Performance
27.0-27.9	16	119	67	51
28.0-28.9	24	127	72	56
29.0-29.9	23	129	73	55
30.0-30.9	20	126	72	54
31.0-31.9	24	119	70	50

TABLE 5

SEX DIFFERENCES ON THE STANFORD-BINET

Test	Male	Female	Difference
1931 L-M			
N	51	59	
M	108.5	115.8	−7.3
SD	14.4	13.3	
*CR**			2.79
1941 L			
N	52	59	
M	107.9	116.1	−8.2
SD	19.0	13.7	
*CR**			2.58
1956 L			
N	52	59	
M	121.6	125.7	−4.1
SD	19.6	10.6	
*CR**			1.36

*A value of 2.57 is significant beyond the .01 level and a value of 1.96 beyond the .05 level.

the first two administrations, this difference was significant beyond the .01 level. On the 1956 testing, it did not reach significance. The males increased their scores from 1941 to 1956 to a greater extent than did the women. For the males, this increase was 13.7 and for the females it was 9.6; the difference of 4.1 is significant beyond the .01 level. Both females and males show a negative correlation between the 1941 Stanford-Binet IQ and the gain in IQ from 1941 to 1956, but the one for females is larger. For men this correlation is −.23 and for women −.67; the difference between these correlations is significant beyond the .01 level. Dividing the male and female groups according to whether they stopped their education at high school graduation or went beyond does not change the findings. The obtained correlations for males are −.28 and −.19 (not significantly different); and for females −.70 and −.65 (not significantly different). The decrease in standard deviation for females from 1941 to 1956 as shown in Table 5 reflects that these findings are due to the tendency of women with IQs which are high at adolescence not to increase their scores by adulthood as much as does the rest of the group.

A study of the individuals scoring high (from 130 to 150 IQ) on the 1956 Stanford-Binet yields further information. Females in this group had always scored high relative to the group: mean IQ of 125.9 in 1931, 127.8 in 1941, and 137.5 in 1956. For the males, however, becoming a member of this group represented a gain from an earlier lower point with most of the gain achieved by adolescence: mean IQ of 115.9 in 1931, 126.2 in 1941, and 140.7 in 1956. (It must be remembered that these 1956 Stanford-Binet IQs are, on the average, 11 points high, since they do not include a correction for the fact that there was a mean IQ increase of 11 points for the total sample.) Slightly but not significantly more women than men in this subgroup stopped education with high school graduation. Of those who started college, not a single man failed to finish, whereas one-third of the women dropped out before finishing. This leaves us with the familiar chicken-egg dilemma. Did the women drop out of college because their relative brightness as compared with men decreased—their promise not being realized in the same way as was that of the men? Or did the earlier termination of their academic education make the intelligence tests administered to them as adults relatively more difficult for them?

An analysis of the relation between IQ at adolescence and continuation of education for the total group was made. There is no significant difference for either sex between mean 1941 IQ of those who stopped their education with graduation from high school and the mean 1941 IQ of those who attended college but did not graduate (males:

100.7 and 102.5; females: 113.8 and 113.0). The mean 1941 IQ, however, of both the 17 males and the 14 females who graduated from college is significantly higher (124.6 and 125.0, respectively) than for those who did not graduate from college. Gain in IQ, on the other hand, is not consistently related to educational attainment. Mean IQ gains for males who did not go beyind high school graduation, who entered college but did not graduate, and who graduated from college were 14.7, 11.7, and 11.8, respectively. Corresponding mean gains for females were 8.2, 12.3, and 9.3, respectively. The differences between the mean gains for either sex are not large enough to be statistically significant with the small number of subjects involved (males: 22, 13, and 17 for successive educational attainments; females: 22, 22, and 14).

Because other investigators of adult intelligence have not reported sex dif-

TABLE 6

ITEMS OF STANFORD-BINET (L) ADULT LEVELS BY MAGNITUDE OF CHANGE IN PERCENT PASSING FROM ADOLESCENCE TO ADULTHOOD

Rank	Location	Item	Percent Passing[a] 1941	1956	Increase	Chi square[b]	Type[c]
1	AA-5	Proverbs I	22.5	81.1	58.6	76.26	RA
2	SAI-1	Vocabulary	35.1	92.8	57.7	80.07	V
3	SAII-1	Vocabulary	13.5	67.6	54.1	67.28	V
4	SAI-6	Essential Similarities	31.5	82.9	51.4	52.42	RA
5	SAI-5	Sentence Building	30.6	77.5	46.9	49.04	V
6.5	AA-6	Ingenuity	30.6	73.9	43.3	38.01	RP
6.5	SAIII-1	Vocabulary	1.8	45.0	43.2	52.84	V
8	AA-3	Difference between Abstract Words	49.5	91.9	42.4	48.05	V
9	SAIII-3	Opposite Analogies	8.1	48.7	40.6	44.89	V
10.5	AA-2	Codes	37.8	77.5	39.7	35.73	RP
10.5	SAII-5	Reconciliation of Opposites	29.7	69.4	39.7	34.89	RA
12	AA-1	Vocabulary	59.4	98.2	38.8	49.41	V
13	SAII-4	Proverbs II	6.3	44.1	37.8	42.81	RA
14	AA-7	Memory for Sentences V	36.9	72.1	35.2	27.63	M
15	SAII-2	Finding Reasons II	14.4	47.8	33.4	28.68	RA
16	AA-8	Reconciliation of Opposites	54.9	86.5	31.6	26.64	RA
17	AA-4	Arithmetic Reasoning	41.4	72.1	30.7	21.22	RP
18	SAIII-2	Orientation: Directions II	14.4	41.4	27.0	20.15	RP
19	SAI-3	Minkus Completion	24.3	50.5	26.2	16.18	V
20	SAI-4	Repeating 6 Digits Reversed	29.7	54.1	24.4	13.45	M
21	SAIII-5	Reasoning	3.6	24.3	20.7	19.82	RP
22	SAII-3	Repeating 8 Digits	17.1	35.1	18.0	9.34	M
23	SAIII-6	Repeating 9 Digits	3.6	16.2	12.6	9.89	M
24	SAIII-4	Paper Cutting II	11.7	22.5	10.8	3.57	RP
25	SAI-2	Enclosed Box Problem	50.4	55.9	5.5	0.62	RP
26	SAII-6	Repeating Thought of Passage	24.3	27.9	3.6	0.09	M

[a] $N = 111$.

[b] A chi square value of 10.83 is significant beyond the .001 level, 6.64 beyond the .01 level, and 3.84 at the .05 level. Hence all but the last three items of this table show a significant increase in percent passing from adolescence to adulthood.

[c] These refer to the classification of Davis and Robinson described previously.

ferences of the nature found here, we cannot look beyond our own data for the answers to some of the questions which are posed. This seems to open up an area worthy of further study designed explicitly to explore this problem.

Increases in Abilities from Adolescence to Adulthood

With the establishment of the fact that there is increase in intelligence scores between adolescence and adulthood, one becomes interested in identifying the kinds of abilities which increase most. Within the range of abilities covered by the adult levels of the Stanford-Binet, it was possible to shed light on this question by comparing the proportion of our group passing each item as adolescents and as adults. The percentages of the present sample passing each item were calculated for the 1941 and 1956 results. Items occurring below a subject's basal age were assumed to have been passed; those occurring above a ceiling were assumed to have been failed. Since nearly all subjects passed in 1956 all those items through Level 14, it was possible to make meaningful comparisons for the four adult levels only. Table 6 lists each of the 26 items at these levels, arranged according to the magnitude of the increase in percent passing over the 15-year period. The final column in the table classifies items by "type," following a system suggested by Davis (1941), into rote memory (M), reasoning (R), and vocabulary (V). The reasoning subscale has been further divided by Robinson[4] into those items which appear to involve mental manipulation of practical, concrete concepts (RP) versus those dealing with more abstract concepts (RA). One might wonder whether those items having the lowest percent passing in 1941 would by artifact show the most increase to 1956. The correlation between percent passing in 1941 and increase from 1941 to 1956 was found, however, to be only .06.

TABLE 7

COMPARISON OF FOUR SUBSCALES ON INCREASE FROM ADOLESCENCE TO ADULTHOOD

Scale	Number of items	Rank order among 26 items		Mean of increases in percent passing
		M	Mdn.	
Vocabulary	8	8.1	7.3	43.7
Reasoning				
Abstract	6	9.9	11.8	42.2
Practical	7	17.4	18.0	25.1
Memory	5	21.0	22.0	18.7

Table 7 presents a comparison of the data which demonstrates the differential increases made by our subjects in the four classifications of items. Immediate rote memory items and those involving practical reasoning showed smaller increases than those involving abstract reasoning and vocabulary. This finding tends to confirm the conclusion of most investigators that growth of functions dependent on continued learning and on the building of concepts is more likely to continue into adulthood than is growth of memory or performance functions.

These data make it clear that one should not present conclusions about the growth of intelligence into the adult years without specifying what kinds of intelligence one means and how one is measuring these. If one administered an intelligence test consisting solely of items measuring vocabulary and abstract reasoning, he would find more increase of "intelligence" into adult-

[4] Nancy M. Robinson made several contributions to the treatment of data for this section of the study.

hood than if he used a test heavily weighted with memory and practical reasoning items. Thus one cannot realistically expect ever to describe the shape of *the* mental growth curve.

A finding in these data which has important implications when it comes to trying to predict amount of subsequent mental growth in an individual adolescent is the negative relationship between level of ability at adolescence and gain in mental age score from adolescence to adulthood. We have found, for example, that the higher the 1941 Stanford-Binet score, whether expressed as IQ or MA, the slightly lower the rate of increase (MA increase/CA increase) to 1956 for the group of 45 subjects who were 14 to 16 years of age in 1941. (For this analysis, subjects less than 14 years of age in 1941 were eliminated in order to minimize the operation of differential rates of mental growth in adolescence.) The correlations expressing this ($r = -.152$ and $-.212$ for IQ and MA, respectively) are not significantly different from zero. Comparable correlations for males and females considered separately are consistent with the findings on sex differences discussed earlier in that the females showed a significant negative relationship, whereas the males did not. The correlations between 1941 IQs and rate of increase to 1956 for 24 males and 21 females were .058 and $-.385$, respectively.

Thus, although the adult scores on intelligence tests can be predicted with some accuracy from the score one obtains as an adolescent (as indicated by the correlations in Table 1), we have not isolated factors which permit us to predict the amount of mental age gain. The fact that adult intelligence can be predicted from adolescent intelligence with such a high degree of accuracy despite the lack of relation between adolescent intelligence and amount of gain to adulthood is, of course, due to the fact that the amount of gain at these levels represents such a relatively small proportion of the total "amount" of intelligence. We can anticipate that there will be an increase in mental age from adolescence to adulthood, but we cannot predict which subjects will gain most. We do know that amount of gain is not dependent upon level of initial mental age, on the extent of education, or on the level of ancestral intelligence (Bradway & Robinson, 1961), in contrast to mental growth in the earlier years where these factors are important (Bradway, 1945a). Our only positive finding in this area is the one pointed out earlier, which is that adolescent girls with high IQs can be expected to increase less in mental age than will girls who test lower, or than males who test at any level.

Prediction of Adult Intelligence

The next questions which arise are those of how early can adult intelligence be predicted, and what are the best measures for predicting it. Or, what are the origins of adult verbal and nonverbal abilities?

An analysis of the predictive value of Stanford-Binet preschool items to adult WAIS scores can help answer these questions. At the time of the 1941 study, the principal investigator devised four subscales from the preschool Form L and Form M Stanford-Binets and computed four subscale quotients for each subject (Bradway, 1945b). These were labeled "verbal quotient," "nonverbal quotient," "memory quotient," and "number concept quotient." Because the Stanford-Binet at the adolescent and adult levels is an almost wholly verbal instrument, it was not possible in that earlier study to determine whether or not one maintains his relative superi-

ority and inferiority in these categories over any length of time. The administration of the Wechsler Adult Intelligence Scale in the present study, however, allows us to make such "profile" comparisons. Corresponding subscale indices were derived from the WAIS by using Comprehension, Similarities, Information, and Vocabulary for comparison with the 1931 verbal quotient; the Performance scale for comparison with the nonverbal quotient; Digit Span with the memory quotient; and Arithmetic with the number concept quotient. Inasmuch as the WAIS results are in standard scores, all that was necessary to obtain indices for these four WAIS categories which were comparable with each other was to divide the verbal total by four and the performance total by five. The mean of these four indices was then obtain for each subject, and his score for each category was expressed as a deviation from his own mean. A similar method was used for the 1931 scores, except that each of

TABLE 8

CORRELATION OF PREDICTIVE PRESCHOOL ITEMS OF FORMS L AND M OF THE STANFORD-BINET WITH WAIS VERBAL QUOTIENTS

Location	Item	r_{bis}	Groups by years of age[a]	Type[b]
M II ½ 6[c]	Obeying Simple Commands	.93***	2½	V
L III 2[c]	Picture Vocabulary	.77**	2½	V
L III 4[c]	Picture Memories	.85***	2½	M
M III 3	Objects by Use	.66*	2½	V
M III 6	Repeating Three Digits	.67*	2½	M
L III ½ 2	Picture Vocabulary	.84**	4	V
L III ½ 5	Objects by Use	.98***	2½	V
L III ½ 6[c]	Comprehension I	.86**	4	V
M III ½ 6	Comprehension I	.74**	2½	V
L IV 1[c]	Picture Vocabulary	.85***	4	V
L IV 2[c]	Objects from Memory	.79**	4	M
L IV 4[c]	Pictorial Identification	.80**	4	V
L IV A[c]	Memory for Sentences I	.80**	4	M
M IV 6	Memory for Sentences I	.70*	2½	M
L IV½ 6	Opposite Analogies I	.73**	4	V
M IV½ 1	Animal Pictures	.64*	4	NV
M IV½ 5	Materials	.74**	4	V
L V 5	Memory for Sentences II	.64*	4	M
M V 2	Number Concept of Three	.74**	4	NC
M V 3[c]	Pictorial Similarities and Differences	.61*	5½	V
L VI 4	Number Concepts	.71**	5½	NC
L VI 5	Pictorial Likenesses and Differences	.72***	5½	V
M VI 2	Copying Bead Chain	.55*	5½	NV
M VI 3[c]	Differences	.60*	5½	V
M VII 4[c]	Three Digits Reversed	.62*	5½	M

[a] There were 17 subjects in the 2½-year-old group; 16 in the 4-year-old; 17 in the 4½-year-old; and 21 in the 5½-year-old. The other groups (2-year-olds, 3-year-olds, 3½-year-olds, and 5-year-olds) were not large enough for valid calculations.

[b] These refer to the Bradway subscales described previously (Bradway, 1945b).

[c] Indicates used in 1960 L-M Revision of the Stanford-Binet.

* Significant beyond .05 level.

** Significant beyond .02 level.

*** Significant beyond .01 level.

the four subscale scores had to be converted into standard scores before the individual profiles of deviation scores could be figured.

The stability of pattern over a period of 25 years was tested by the correlation between the deviation scores calculated as each individual's deviation from his own mean for the 1931 subscales, and the deviation scores also calculated as each individual's deviation from his own mean for the 1956 WAIS. Three of these four correlations are significant in the positive direction. For the verbal component, the Pearsonian r is .236 significant at the .02 level; for the nonverbal component, .191 significant at the .05 level; and for the memory component, .301 significant at the .01 level. For the number concept component,

TABLE 9

CORRELATIONS OF PREDICTIVE PRESCHOOL ITEMS OF FORMS L AND M OF THE STANFORD-BINET WITH WAIS PERFORMANCE QUOTIENTS

Location	Item	r_{bis}	Groups by years of age[a]	Type[b]
M III 2[c]	Picture Vocabulary	.90***	2½	V
L III½ 6[c]	Comprehension	.75*	4	V
L IV 1[c]	Picture Vocabulary	.75*	4	V
L IV 2[c]	Objects from Memory	.71*	4	M
L IV 4[c]	Pictorial Identification	.87***	4	V
L IV 5[c]	Discrimination of Forms	.71*	4	NV
M IV 3[c]	Opposite Analogies I	.70*	4	V
M IV 4[c]	Pictorial Identification	.70*	2½	V
M IV 6	Memory for Sentences I	.75*	4	M
L IV½ 4	Materials	.68*	4½	V
L IV½ A	Pictorial Identification	.71*	4	V
M IV½ 1	Animal Pictures	.63*	4	NV
M IV½ 3	Repeating Four Digits	.70*	2½	M
M IV½ 5[c]	Materials	.69*	4	V
L V 4[c]	Copying a Square	.74*	4	NV
L V 6	Counting Four Objects	−.62*	4½	NC
M V 2	Number Concept of Three	.75**	4	NC
M V 5	Comprehension II	.78***	5½	V
L VI 1	Vocabulary	.86***	5½	V
L VI 2	Bead Chain from Memory	.68*	4½	M
		.58*	5½	
L VI 3[c]	Mutilated Pictures	.82**	4½	NV
		.77***	5½	
L VI 5	Pictorial Likenesses and Differences	.53*	5½	V
M VI 2	Copying Bead Chain	.94***	5½	NV
M VI 5	Counting 13 Pennies	.67**	5½	NC
M VI 6	Opposite Analogies I	.73**	5½	V
L VII 1[c]	Picture Absurdities I	.61*	5½	V
M VII 4	Three Digits Reversed	.60*	5½	M

[a] There were 17 subjects in the 2½-year-old group, 16 in the 4-year-old, 17 in the 4½-year-old, and 21 in the 5½-year-old. The other groups (2-year-olds, 3-year-olds, 3½-year-olds, and 5-year-olds) were not large enough for valid calculations.

[b] These refer to the Bradway subscales described previously (Bradway, 1945b).

[c] Indicates used in 1960 L-M Revision of the Stanford-Binet.

* Significant beyond .05 level.

** Significant beyond .02 level.

*** Significant beyond .01 level.

(figured on 79 subjects rather than 110 because for 31 subjects no such ratio could be computed at the preschool ages) the correlation was only .089, not significant, suggesting that relative strength or weakness in this ability is not stable over a 25-year period. Inasmuch, however, as one's score on this scale contributes to his deviation scores on the other three, further analysis is necessary before drawing such a conclusion.

To study the predictive value of preschool items in more detail, biserial correlations of each Stanford-Binet preschool item to adult WAIS verbal quotient and adult WAIS performance quotient were calculated. This analysis is summarized in Tables 8 and 9 in which are listed those items whose biserial *r*'s were significant beyond the .05 level for the indicated age groups.

The figures in the column headed "Groups by years of age" indicate the age group for which the corresponding biserial *r* was figured. Since there was no way of combining the age groups, biserial *r*'s were figured for each 1931 age group for which an item was pertinent, which was defined for purposes of this analysis as each age group in which the number of subjects was 15 or more and for which the percent passing the item was between 25 and 75. The distribution by type of the total 79 items for which biserial *r*'s at one or more age levels were computed were as follows: 37 verbal, 24 nonverbal, 12 memory, and 6 number concept. The type designation follows that previously described by Bradway (1945b). Comparison of the frequency of each type found to have significant predictive value at one or more age levels and the total number of such items studied shows the following: of 37 verbal items, 14 (38%) predicted WAIS verbal IQ and 14 (38%) predicted WAIS performance IQ; of the 24 nonverbal items, 2 (8%) predicted WAIS verbal IQ and 5 (21%) predicted WAIS performance IQ; of the 12 memory items, 7 (59%) predicted WAIS verbal IQ and 5 (41%) predicted WAIS performance IQ; of the 6 number concept items, 2 (33%) predicted WAIS verbal IQ and 3 (50%) predicted WAIS performance IQ. This kind of analysis is only suggestive because the number of items is small, but it supplements in more detail the findings for the subscales reported above. In summary: verbal items predict both verbal and performance adult ability better than do nonverbal items; each kind of memory item (memory for digits forward and reversed, for sentences, for objects, for pictures, and for bead chain) predicts both WAIS verbal and performance ability; number concept items are inconsistent in their predictive value —the only significant negative correlation obtained being for a number concept item.

When one compares these results with those for the subscale deviation scores reported above, similarities are found. Both analyses suggest the greater value of verbal and memory items over nonverbal items in predicting adult abilities, and the unreliability of number concept items at the preschool level as predictors of adult abilities.

Measurement of Intelligence

After reviewing these data we are in a better position to approach the problems of what index or indices to use in expressing adult intelligence and how best to predict it. If the 1937 Stanford-Binet is used to measure adult intelligence, a use for which it was not specifically designed, it is apparent that a correctional factor of some kind must be applied to IQs obtained according to the published instructions. Merrill kindly applied the deviation scores provided in the 1960 Revision (Terman & Merrill,

1960) to our data. The resulting intelligence quotients are 109.3 in 1931, 109.4 in 1941, and 119.4 in 1956. These findings of identity between preschool and adolescence and a 10 point rise to adulthood indicate that the correction inherent in the 1960 deviation scores does not solve the problem. For our research purposes we rejected the use of a multiplicative factor because this would have changed the standard deviations, and used instead a subtractive factor which, since the standard deviations were nearly identical, was sufficient to equate the adult set of scores with the two previous ones. This, although a satisfactory solution for research purposes, may not be the optimal solution from the point of view of the clinician. Perhaps the possibility of revised deviation tables, accounting for greater increase into adulthood than do those published with the 1960 Revision, should be explored.

This problem is not encountered in measuring adult intelligence with the WAIS which was specifically designed for the measurement of adult intelligence. Moreover, our data demonstrate that there are different kinds of ability which have a degree of constancy throughout life. The structure of the WAIS is a recognition of this.

When new tests are constructed, those for adults would do well to provide valid measurements of different kinds of abilities, because the practicing clinician needs such information to evaluate selective impairment of intellectual functioning and for vocational guidance and occupational placement. In contrast, a preschool test heavily weighted with verbal items should be a more valid predictor of adult abilities, both verbal and nonverbal, than one utilizing many performance-type items. Moreover, the former would be consistent with the usual purpose for which preschool tests are used in that the prediction of school performance puts a heavy demand on verbal abilities.

The present data demonstrate that the defining of adult intelligence is complex. It is not expressed by a single index in terms of abilities, in terms of one's place relative to others of the same age group, or of one's location on a scale of development and decline. As Cameron and Magaret (1951, p. 158) have said in alluding to the use of tests in defining behavior, "Performance on test items alone can no more define biosocial immaturity than can a refractive index define the complex process of seeing."

SUMMARY

A group of 111 subjects who had been tested with the Stanford-Binet (S-B) at the preschool and adolescent levels were administered the S-B and WAIS at adulthood (25 years after initial testing). Correlations of preschool IQs with adult S-B and full WAIS IQs are .59 and .64. Correlations of adolescent IQs with adult S-B and full WAIS IQs are .85 and .80. The S-B IQ shows a mean increase of 11 points from adolescence to adulthood, which corresponds to a mental age gain of 20 months beyond what would be expected if mental growth ceases at 16 years. The data were analyzed in an attempt to determine the point at which mental growth, as measured by these tests, ceases, but it could not be determined.

The males show significantly more IQ gain from adolescence to adulthood than do the females ($p < .01$). Factors associated with amount of gain were not isolated except that girls with high IQs at adolescence increase less than do girls who initially test lower or than males who test at any level. There is no evidence that educational attainment is related to amount of gain.

Analysis of increases in percent passing S-B items from adolescence to adulthood indicates that intellectual growth after adolescence is greater in abstract reasoning and vocabulary than in rote memory and practical reasoning.

Analysis of the value of S-B preschool items in predicting adult WAIS scores shows some stability of pattern of individual differences in verbal versus nonverbal versus memory superiority over the 25 years. It also shows the greater value of verbal and memory items over nonverbal items in predicting both verbal and performance adult IQs, and the unreliability of number concept items as predictors of adult abilities.

The data are interpreted as demonstrating that adult intelligence is complex and cannot be meaningfully expressed by a single index.

REFERENCES

BAYLEY, NANCY. On the growth of intelligence. *Amer. Psychologist*, 1955, **10**, 805-818.

BAYLEY, NANCY. Data on the growth of intelligence between 16 and 21 years as measured by the Wechsler-Bellevue scale. *J. genet. Psychol.*, 1957, **90**, 3-15.

BAYLEY, NANCY, & ODEN, MELITA H. The maintenance of intellectual ability in gifted adults. *J. Gerontol.*, 1955, **10**, 91-107.

BRADWAY, KATHERINE P. IQ constancy on the Revised Stanford-Binet from the preschool to the junior high school level. *J. genet. Psychol.*, 1944, **65**, 197-217.

BRADWAY, KATHERINE P. An experimental study of factors associated with Stanford-Binet IQ changes from the preschool to the junior high school. *J. genet. Psychol.*, 1945, **66**, 107-128. (a)

BRADWAY, KATHERINE P. Predictive value of Stanford-Binet preschool items. *J. educ. Psychol.*, 1945, **36**, 1-16. (b)

BRADWAY, KATHERINE P., & ROBINSON, NANCY M. Significant IQ changes in twenty-five years: A follow-up. *J. educ. Psychol.*, 1961, **52**, 74-79.

BRADWAY, KATHERINE P., THOMPSON, CLARE W., & CRAVENS, R. B. Preschool IQs after twenty-five years. *J. educ. Psychol.*, 1958, **49**, 278-281.

CAMERON, N., & MAGARET, ANN. *Behavior pathology.* Boston: Houghton Mifflin, 1951.

DAVIS, F. B. The derivation of three subscores from the 1937 Revision of the Stanford-Binet Scales. *J. consult. Psychol.*, 1941, **5**, 287-291.

INHELDER, B., & PIAGET, J. *The growth of logical thinking from childhood to adolescence.* New York: Basic Books, 1958.

MINER, J. B. *Intelligence in the United States.* New York: Springer, 1957.

OWENS, W. A. Age and mental abilities: A longitudinal study. *Genet. psychol. Monogr.*, 1953, **48**, 3-54.

TERMAN, L. M., & MERRILL, MAUD A. *Stanford-Binet Intelligence Scale.* Boston: Houghton Mifflin, 1960.

(Received July 17, 1961)

Journal of Educational Psychology
1962, Vol. 53, No. 1, 15-26

FOREIGN LANGUAGE LEARNING ABILITY[1]

PAUL PIMSLEUR
Ohio State University

ROBERT P. STOCKWELL, AND ANDREW L. COMREY
University of California, Los Angeles

This report on 2 studies of factors involved in learning French is based on 23 tests administered to 208 students in college French, and on 22 tests administered to 202 additional students a year later. The factors extracted and rotated in each study included: Verbal, Reasoning, Speed of Articulation, Pitch Discrimination, Timbre Discrimination, Interest, and Biographic. Multiple correlation test selection analyses resulted in (a) R = .65, using 6 tests to predict Cooperative French Test scores; (b) R = .41, using 5 tests to predict aural comprehension; and (c) R = .41, using 5 tests to predict speaking proficiency. Verbal IQ and Interest (motivation) appear to be the most important factors in college foreign language learning; Reasoning, Word Fluency, and Pitch Discrimination also contribute.

It is popularly believed that foreign languages are analogous to mathematics or music in that a special talent is required to learn any of them. This belief appears fallacious and even harmful in that it may adversely affect the achievement of those who think that they do not possess this special talent. The investigation of which this paper is a part seeks to reduce the so-called "talent for languages" to a set of well-defined, measurable components. Tests of the components will then be used to predict probable success in foreign language courses. The attempt to isolate and describe as many as possible of the components affecting foreign language learning should lead to improved understanding of the learning processes involved. Presumably, the tests developed for predictive purposes will also be found useful in diagnosing the sorts of difficulties encountered by poor language learners.

As Carroll's (1960) thorough review of the literature of foreign language prognosis indicates, considerable success has been achieved in predicting results of intensive language courses, such as those given in the military services. However, attempts to predict achievement in high school and college courses have generally been less successful. There appear to be influences present in the typical school course which do not enter into intensive courses, and which need to be taken into account to attain high predictive validity. Prognostic tests composed entirely of intellectual tasks probably owe their rather low validity to their failure to include important nonintellectual variables. Motivation factors almost certainly affect achievement, and personality may also be important. Another kind of prognostic test, the work sample, sometimes is a good predictive device, but sheds no light on the language learning process. The studies reported in this paper attempt to deal with a number of fairly precise intellectual and motivational factors. Personality factors, not included here, will be introduced in subsequent research. The previous research most closely related to that reported here is the work of Carroll (1958).

The two studies which are the subject

[1]Since June 1960, the research reported herein has been performed pursuant to a contract with the Office of Education, United States Department of Health, Education, and Welfare.

of this paper were conducted at UCLA in the spring semesters of 1958 and 1959, respectively. In both cases, the subjects were students in second semester French courses, 208 in the first study, and 202 in the second. The two samples appear closely comparable, and quite typical of a moderately good group of American undergraduates. Their average age was 19 years; most were graduates of California public high schools who gained admission to college by achieving a B average; most were taking French to fulfill degree requirements.

The data consist of scores on a number of predictor variables and several criteria. In the light of the current shift toward teaching the spoken language, it was of particular interest to attempt differential prediction of traditional (grammar-reading) achievement on the one hand, and oral-aural achievement on the other.

All tests were administered to groups, under good testing conditions during regular class hours. Tests requiring speaking or listening were administered in the language laboratory. Tests similar in nature (e.g., Reading Aloud I and II) were given at least a week apart. The predictor tests were given early in the semester, the criterion tests at the end.

The two studies will be reported in the following order:

Study I—Variables, factor analysis, multiple correlation analysis

Study II—Variables, factor analysis, multiple correlation analysis

Results

STUDY I

Variables

The data for the first study consisted of scores on 21 test variables and two criteria. The test variables were chosen to represent factors which, on the basis of teaching experience and linguistic theory, were hypothesized to be related to success in learning a foreign language. The hypothesized factors were associative memory, analytic reasoning, reasoning by analogy, physical dexterity in articulation, ability to change linguistic set, auditory discrimination, interest in foreign languages, and several others. Tests were either found or constructed to represent the hypothesized factors in as "pure" a manner as possible.

It appeared desirable to try out a large number of such predictor variables. As there were more tests than could be given in a single semester, the first study omitted several factors of potential importance, such as auditory discrimination and language interest, which then appeared in the second study.

The 23 variables which provided the data for the first study are as follows:

1. Modern Language Aptitude Test (MLAT), Part III—Spelling Clues (Carroll & Sapon, 1955). The subject chooses which of five words has the same meaning as the word represented in abbreviated form. Sample: kataklzm—(*a*) mountain lion, (*b*) disaster, (*c*) sheep, (*d*) chemical reagent, (*e*) population. Highly speeded: 50 items, 8 minutes.

2. MLAT, Part I—Number Learning. By tape recording, the subject is taught a simple artificial system of number expression utilizing nonsense syllables. He is then asked to write down the arabic numeral equivalents of a list of 2- and 3-digit numbers in the artificial system spoken at a fairly rapid pace on the tape. This version utilizes the digits 0, 1, 2, 3, 4. Fifteen items, total time 10 minutes.

3. MLAT, Part IV—Words in Sentences. This test was designed to measure the ability to understand the function of words and phrases in sentence structure, without calling on knowledge of grammatical terminology. Each item consists of a "key sentence" with a word or phrase capitalized, followed by one or more sentences with other words or phrases underlined and numbered. The subject is directed to pick the word or phrase in the second sentence (or sentence-group)

which does the same thing in that sentence as the capitalized word does in the key sentence.

Sample:

He spoke VERY well of you.
Suddenly the music became *quite loud.*
1 2 3 4

The test contains 45 items. The time allowed is 15 minutes.

4. MLAT, Part V—Paired Associates. The subject studies a list of 24 "Kurdish-English" vocabulary equivalents for 2 minutes; in the next 2 minutes he practices recalling the English meanings, and he then has 4 minutes to complete from memory a multiple-choice test of the presented vocabulary (24 items).

5. Letter Series (adapted from Guilford). A series of seven letters is given, to which the subject must add the next two. Sample:

D E F D E F – – .

6. Reading Aloud I (speed). The subject is given time to study a meaningful paragraph of English prose, then is asked to record his reading of it, "speaking as quickly as possible while still remaining intelligible." Score is number of words read in 30 seconds.

7. Same test (accuracy). Number of errors made in 6.

8. Reading Aloud II (speed). The subject is given time to study a meaningless paragraph made up of English words put together at random. He then records his reading of it "speaking as quickly as possible while still remaining intelligible." Score is number of words read in 30 seconds.

9. Same test (accuracy). Number of errors made in 8.

10. Paraphrase. The subject is to give as many paraphrases as possible for a given phrase. Sample: for TO DIE, the subjects might give "to kick the bucket," "to cash in one's chips," etc. Score is number of paraphrases given in 4 minutes.

11. Rhymes. The subject is to give as many words as possible that rhyme with four given words (LAKE, CLOUD, SO, GRASS). Score is number of rhymes given in 2 minutes.

12. Synonyms. The subject is to give as many synonyms as possible for four given words. Sample: for GO, the subject might give "ride," "drive," "fly," "travel," etc. Score is number of synonyms given in 4 minutes.

13. Phonetic Perception. The subject hears three sounds and must tell which one is different from the other two. The test uses distinctions which are nonphonemic in English (e.g., "r" vs. "rr"). Twenty items, 5 minutes.

14. Linguistic Analysis I. The subject is given a list of 22 foreign forms (adapted from Kabardian) together with their English equivalents. By consulting the given forms, he is to deduce how other things are said in this language. Ten multiple-choice items, 12 minutes.

15. Linguistic Analysis II. The subject is given a list of 19 forms in a foreign language (adapted from Kabardian) together with their English equivalents. By consulting the given forms, he is to deduce the meanings of a number of other foreign forms given to him. Ten multiple-choice items, 12 minutes.

16. Verbal Comprehension (Guilford-Zimmerman). A test of English vocabulary knowledge. Forty items, 6 minutes.

17. Age. The subject gives his age at last birthday and is placed in one of five categories: under 18, 18-20, 20-22, 22-25, over 25.

18. Sex. Female or male, scored zero or one.

19. Bilingualism. To be rated as "bilingual" a subject has to answer either or both of the following questions affirmatively: Is a language other than English spoken regularly in your home? Is your native language other than English? Scored zero for mono-, one for bilingualism.

20. High School Language Grades. An average of grades in high school language courses.

21. High School Math-Science Grades. An average of grades in math and science courses in high school.

22. (Criterion) French II Final Grades. Letter grade assigned by teacher at end of second semester of college French. (A, B, C, D, F = 4, 3, 2, 1, 0).

23. (Criterion) French Speaking Proficiency Test. A five-part recorded test of ability to speak French at end of second semester. Scored by native judges.

Factor Analysis

A matrix of zero-order Pearson correlations was computed from the scores of the 208 subjects on the 23 variables just described. These correlations, together with the means and standard deviations of the variables, are shown

in Table A.[2] From this matrix, eight centroid factors were extracted and then rotated, using Kaiser's normal varimax criterion (Kaiser, 1958).[3] Table B presents the rotated factor matrix. Loadings of less than .30 were not considered significant. Descriptions of the eight factors follow:

Factor A: Reasoning. This factor is made up of seven variables (loadings shown at right).

2.	Number Learning	.57
21.	High School Math-Science Grades	.50
14.	Linguistic Analysis I	.48
3.	Words in Sentences	.43
15.	Linguistic Analysis II	.35
5.	Letter Series	.34
22.	French II Grades (criterion)	.34

This factor resembles others variously called General Reasoning, Verbal Reasoning, Induction, etc., in previous factor analytic findings. It appears to include the ability to induce the orderly principle in a group of examples (e.g., Letter Series), and to apply this principle to a variety of new examples (e.g., Number Learning).

Factor B: Word Fluency. The variables with sizable loadings on this factor are:

10.	Paraphrase	.65
12.	Synonyms	.64
11.	Rhymes	.41

[2]Tables A, B, C, and D have been deposited with the American Documentation Institute. Order Document No. 6943 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

[3]The factor analytic computations in Study I were performed on SWAC, an electronic computer located on the campus of UCLA and supported by the Office of Naval Research.

The identification as Word Fluency requires some amplification. Word Fluency is usually defined as the ability to produce words which conform to certain structural limitations, e.g., four-letter words of the types s – – e. In the present case, the limitations are not structural but semantic. Both kinds of tests entered into factors called Word Fluency by previous investigators (Carroll, 1941; Thurstone, 1941). The factor found here most nearly resembles Guilford and Christensen's (1956) Eduction of Conceptual Correlates. The name Word Fluency has been retained because it is more familiar.

Factor C: Biographic. This factor is made up of:

10.	Sex—scoring code: feminine = 0, masculine = 1	.64
20.	High School Language Grades	−.58
17.	Age	.52
4.	Paired Associates	−.34
3.	Words in Sentences	−.31

This factor seems to reiterate the well-known fact that girls do better than boys in high school language courses. However, at the college level, sex fails to correlate significantly with either criterion (−.02 and −.08, respectively).

Factor D: Achievement in French. This factor is made up of the two criteria:

23.	French Speaking Test	−.65
22.	French II Grades	−.57

It seems to represent a unique portion of the variance in the criteria. It may be pointed out that the two criteria were arrived at in different ways, by different judges.

Factor E: Speed of Articulation. Two variables have significant loadings:

8.	Reading Aloud II (speed)	−.71
6.	Reading Aloud I (speed)	−.68

The identification seems clear.

Factor F: ? The variables included in this factor are:

5.	Letter Series	−.42
13.	Phonetic Perception	−.37
19.	Bilingualism	.34
15.	Linguistic Analysis I	−.32

The meaning of this factor is not clear, beyond the fact the bilinguals perform poorly in the three other variables. No interpretation will be attempted, as these variables appear again in the next study.

Factor G: Response Set. The two variables composing this factor are:

9.	Reading Aloud II (errors)	−.48
7.	Reading Aloud I (errors)	−.44

This factor, reflecting the number of errors made while reading a passage at top speed, may be interpreted as Response Set, i.e., the tendency of some to avoid mistakes by working a little slower, while others work very rapidly even at the risk of making mistakes.

Factor H: Verbal. Three variables make up this factor:

16.	Verbal Comprehension	.59
1.	Spelling Clues	.43
11.	Rhymes	.37

The chief loading being that of an English vocabulary test, this may be identified as the Verbal or Verbal Knowledge factor.

Summarizing the factors whose identification appears clear, they are: Reasoning, Word Fluency, Biographic, Achievement in French, Speed of Articulation, Response Set, and Verbal.

Multiple Correlation Analysis

A multiple correlation coefficient, *R*, was computed[4] for each of the two criteria, using as predictors the entire battery of 21 tests. The results, together with the proportion of variance contributed by each variable, are presented in Table 1, Columns A and C.

Reduction was attempted in the size of the batteries. The three or four tests making the highest contributions were chosen as a basic battery. To these were added other tests, singly or several at a time, in many different combinations, and the resulting *R*s computed. The batteries shown in Table 1, Columns B and D, are those which minimize the number of tests while maximizing the multiple correlation coefficient.

For predicting French grades, the minimum battery consists of Number Learning (or Spelling Clues), Words in Sentences, Letter Series, Reading Aloud I, Paraphrase, Linguistic Analysis II, age, and high school math-science grades. This seven-test battery yields a multiple *R* of .43. The inclusion of Number Learning or Spelling Clues gives approximately the same results; the inclusion of both does not raise the *R* appreciably.

For predicting Speaking Test scores, the minimum battery consists of Spelling Clues, Letter Series, Reading Aloud II, Verbal Comprehension, and Bilingualism. This five-test battery yields a multiple *R* of .42.

In evaluating the multiple correlation coefficients, several factors which militate against accurate prediction must be taken into account. The first of these is the unreliability of the teacher grades used as one of the criteria. If grades are assumed to have a reliability of, say,

[4] This part of the analysis used the BIMD-6 program supplied by the Biostatistics Laboratory, Department of Preventive Medicine, UCLA. The computations were performed on an IBM 709 computer, through the cooperation of the Western Data Processing Center, UCLA.

TABLE 1

STUDY I: MULTIPLE CORRELATION ANALYSIS

Variable	French II Grades		French Speaking Test	
	A (21 tests)	B (7 tests)	C (21 tests)	D (5 tests)
1. Spelling Clues	.033		.058	.058
2. Number Learning	.025	.036	.002	
3. Words in Sentences	.047	.063	.007	
4. Paired Associates	.002		.003	
5. Letter Series	.008		.019	.024
6. Reading Aloud I (speed)	.021	.029	.023	
7. Reading Aloud I (errors)	.000		.003	
8. Reading Aloud II (speed)	.002		.010	.027
9. Reading Aloud II (errors)	.014		.001	
10. Paraphrase	.009	.012	.001	
11. Rhymes	.001		.000	
12. Synonyms	.000		.000	
13. Phonetic Perception	.005		.008	
14. Linguistic Analysis I	.000		.000	
15. Linguistic Analysis II	.010	.012	.004	
16. Verbal Comprehension	.002		.036	.044
17. Age	.012	.012	.002	
18. Sex (feminine = 0, masculine = 1)	.001		.001	
19. Bilingualism	.022	.024	.028	.023
20. High School Language Grades	.004		.000	
21. High School Math-Science Grades	.008		.001	
	$R = .478$	$R = .433$	$R = .454$	$R = .418$

Note.—Proportion of variance contributed by each variable to the prediction of each criterion: A and C —using all tests, B and D—using the most economical battery.

.70, then the correction for attenuation raises the multiple R from .43 to .52. The Speaking Test criterion might be similarly corrected, though there is some evidence to show that it has high reliability (Pimsleur, 1961). A second consideration is the fact that certain potentially important variables were not included. The use of additional variables and more reliable criteria are features of the second study. A third consideration is the restriction in range of ability among the examinees. The college population is highly select; the second semester population of French students is even more so. This homogeneity makes accurate prediction very difficult.

STUDY II

Variables

In Study II, a somewhat different set of variables was used. Those which had proven uninteresting in Study I were discarded. Others were added to tap additional factors, e.g., Numbers 9 through 12, dealing with an auditory factor, and Numbers 13 and 14, treating interest or motivation. Despite their good results in the first study, the MLAT parts were dropped because they contain copyright material and so could not be used in developing a new battery.

The last three variables are three

criteria, all new to this study. The Cooperative French Test was used as a measure of achievement in the reading and writing goals. Oral achievement was rated by a laboratory instructor who had listened to students and graded them in the lab once a week for a full semester. Achievement in understanding spoken French was measured by the Pictorial Auditory Comprehension Test, an objectively scored test.

The variables for Study II are these:

1. Verbal Comprehension (Guilford-Zimmerman). See Study I for description.
2. Linguistic Analysis I. See Study I.
3. Ship Destination Test (Christensen-Guilford). Find the distance from ship to shore considering the influence of several variables. Thirty-three items, 15 minutes.
4. Grammatical Transformations. The subject is given a sample transformation (e.g., "They are good. Books are interesting." → "Good books are interesting.") In a new item, he is to select from among four choices the one which represents a transformation similar to the sample (e.g., "It is noisy. Telephones are annoying." → "Noisy telephones are annoying.") Twenty items, 9 minutes.
5. Paraphrase. See Study I.
6. Rhymes. See Study I.
7. Reading Aloud I. See Study I.
8. Reading Aloud II. See Study I.
9. Phonetic Perception. See Study I.
10. Chinese Pitch Perception. Subject is taught three Chinese words which differ only in pitch. These are then imbedded in Chinese sentences and subject must tell which he hears. Thirty items, 10 minutes.
11. Seashore Pitch Test. Subject hears two tones, must tell whether the second is higher or lower than the first. Thirty-five items, 5 minutes.
12. Seashore Timbre Test. Subject hears two chords, must tell whether the second is the same or different than the first. Thirty-five items, 5 minutes.
13. Interest I. Twenty items relating to interest in language. Subject answers each on a five-point scale.
14. Interest II. The subject is asked to rate on a five-point scale the extent of his interest in studying the foreign language he is now studying.
15. Bilingualism. See Study I.
16. Age. See Study I.
17. Sex. See Study I.
18. High School Language Grades. See Study I.
19. High School Math-Science Grades. See Study I.
20. (Criterion) Cooperative French Test, Advanced Forms Q and R. A standardized test of achievement in reading, grammar, and vocabulary. Total scores were used.
21. (Criterion) Lab Grade. An estimate of speaking ability given by laboratory instructor on basis of listening to student all semester, plus a final oral test. Grades from 0 to 11.
22. (Criterion) Pictorial Auditory Comprehension Test. An objectively scored test of French listening comprehension. Subject must select from among four pictures the one which correctly illustrates the sentence he has just heard. Tape recorded, 50 items, 20 minutes.

Factor Analysis

From the scores of 202 subjects on the 22 variables a matrix of zero-order correlations was computed. It is presented, together with the means and standard deviations of the variables, in Table C. From this, eight factors were extracted and rotated.[5] The rotated factor matrix is presented in Table D.

The analysis yielded eight factors, all of which could be identified, though with varying degrees of assurance.

Factor A: Achievement in French (Verbal). The first factor is composed of seven variables (loadings at right).

20.	Cooperative French Test (criterion)	.73
21.	Lab Oral Grade (criterion)	.55
22.	Aural Test (criterion)	.52
1.	Verbal Comprehension	.47
18.	High School Language Grades	.44
13.	Interest I	.35
17.	Sex	−.31

This factor is ambiguous. The first three variables are the three criteria, and

[5] This part of the analysis used the BIMD-17 program, on the 709 computer.

appear therefore to compose a factor specific to achievement in learning French. Here, they are associated however with a verbal component (Tests 1 and 18), and a component of motivation or interest (Test 13).

Factor B: Speed of Articulation. This factor is made up of:

8. Reading Aloud II	.77
7. Reading Aloud I	.71

The identification is clear. However, these variables did not correlate highly with the criteria, as they did in the first study. For example, Reading Aloud I correlated .20 (significant beyond the .01 level) with the first criterion in the previous study, but only .04 (nonsignificant) with the first criterion in the present study.

Factor C: Reasoning. The components of this factor are:

19. High School Math-Science Grades	.51
3. Ship Destination Test	.46
2. Linguistic Analysis I	.37
4. Grammatical Transformations	.32

This factor seems easily identified as Reasoning, or General Reasoning. The first two tests (19 and 3) are nonverbal in nature, while the other two (2 and 4) involve linguistic tasks. This factor shows a close relation between verbal and nonverbal reasoning.

Factor D: Biographic. This factor consists of:

16. Age	.56
1. Verbal Comprehension	.39
17. Sex	.37
18. High School Language Grades	−.31

It seems appropriate to consider this a biographic factor, though its interpretation is not clear. It is of interest that sex, which did not correlate significantly with the criteria in the first study, did do so in the second.

Factor E: Pitch Discrimination. The variables which form this factor are:

11. Seashore Pitch Test	.50
10. Chinese Pitch Test	.44
4. Grammatical Transformations	.37

The identification as pitch discrimination seems correct. It should be noted that the two pitch tests (11 and 10) are quite different in nature, the former involving pure tones and the latter involving language. These two pitch tests correlate significantly with the first and third criteria, but not with the second one. The presence of Test 4 on this factor is not readily explainable.

Factor F: Word Fluency. This factor is made up of:

5. Paraphrase	.50
6. Rhymes	.41
15. Bilingualism	−.38

The identification is clear. The presence of Test 15 merely means that foreign subjects have less word fluency in English than do Americans.

Factor G: Interest in Languages. The two variables composing this factor are:

14. Interest II	.70
13. Interest I	.67

The identification presents no difficulty. Both these variables correlate significantly with all three criteria.

Factor H: Timbre Discrimination. This factor consists of:

9. Phonetic Perception	.36
12. Seashore Timbre Test	.35

The first of these variables (9) involves actual language sounds, while the other

(12) involves nonlanguage tones. Though the loadings are low here, it is nevertheless intriguing that pitch and timbre appear as two distinct factors. These two variables correlate significantly with the third (aural) criterion, but not with the other two.

Multiple Correlation Analysis

The test selection procedure differs somewhat from that used in the first study. Here, a stepwise regression program was used in conjunction with a multiple correlation program, the former to find the best one-test battery, best two-test battery, etc., up to 9 or 10 tests, and the latter to determine the multiple R and the contributions to variance of each test in each of the best batteries.[6]

Results of the analysis are presented in Table 2. As can be seen in Column A of Table 2, the criterion of Cooperative French Test scores can be predicted to the extent of $R = .673$ when all 19 tests are used. The number of tests may be reduced considerably, without much loss in predictive accuracy. Column B shows the best six-test battery for predicting this criterion. This is a modest battery which can be administered in a 50-minute class period; it yielded an R of

[6]The programs referred to are BIMD-6 and BIMD-9.

TABLE 2

STUDY II: MULTIPLE CORRELATION ANALYSIS

Variable	Criterion 20: Cooperative French Test		Criterion 21: Lab (oral) Grades		Criterion 22: Auditory Comprehension	
	A (19 tests)	B (6 tests)	C (19 tests)	D (5 tests)	E (19 tests)	F (5 tests)
1. Verbal Comprehension	.193	.193	.021	.021	.075	.075
2. Linguistic Analysis	.017		.001		.030	.030
3. Ship Destination	.000		.000		.000	
4. Grammatical Transformations	.004		.002		.001	
5. Paraphrase	.000		.000		.004	
6. Rhymes	.010		.015		.003	
7. Reading Aloud I	.002		.006		.000	
8. Reading Aloud II	.001		.018		.001	
9. Phonetic Perception	.001		.002		.008	
10. Chinese Pitch	.019	.024	.005		.011	.015
11. Seashore Pitch	.000		.000		.002	
12. Seashore Timbre	.004		.000		.009	.013
13. Interest I	.123	.125	.040		.032	.031
14. Interest II	.020	.021	.009	.052	.001	
15. Bilingualism	.007		.003	.009	.000	
16. Age	.006		.008		.000	
17. Sex	.022	.037	.013		.003	
18. High School Language	.019	.024	.048	.073	.005	
19. High School Math-Science	.003		.016	.013	.005	
	$R = .673$	$R = .652$	$R = .457$	$R = .410$	$R = .436$	$R = .405$

Note.—Proportion of variance contributed by each variable to the prediction of each of 3 criteria: A,C,E—using all tests; B,D,F—using the most economical battery.

.652 (estimated shrinkage, $R = .646$).[7] Two improvements in the second study as compared with the first probably account for the more satisfactory result. They are the use of a more reliable criterion, i.e., a standardized test in place of teacher grades, and the inclusion of new variables which turned out to be quite important.

The other two criteria can also be predicted by a small battery of tests, as may be seen in Columns D and F of Table 2. However, these multiple Rs are considerably lower than those for the Cooperative French Test, probably due to the lower reliability of the other two criteria.

RESULTS

Prognosis

By far the best predictive accuracy was obtained in the prediction of Cooperative French Test scores, in Study II. A multiple correlation coefficient of $R = .652$ was arrived at by a six-test battery that can be administered in less than an hour. While this battery has not yet been validated on other samples (such work is under way), it appears to be as good a device for predicting success in foreign language learning in college as is now available. The regression equation for predicting Cooperative French Test scores is:

$$Y' = 47.37 + .47X_1 + .18X_{10} + .09X_{13} + 1.21X_{14} - 2.15X_{17} + 1.82X_{18} \quad [1]$$

As for the other two criteria, speaking ability and listening comprehension, the multiple correlation coefficient obtained for each of them was approximately .41, with a battery of five tests. The regression equations associated with the prediction of these two criteria in Study II are, respectively:

$$Y' = 3.25 + .04X_1 + .31X_{14} + .64X_{15} + .98X_{18} - .37X_{19} \quad [2]$$

$$Y' = 17.80 + .25X_1 + .37X_2 + .13X_{10} + .16X_{12} + .09X_{13} \quad [3]$$

Diagnosis

The manner in which prediction of Cooperative French Test scores was achieved merits attention. The tests contributing to this prediction are Verbal Comprehension, Interest I and II, Chinese Pitch, Sex, and High School Language Grades; i.e., they represent factors of verbal intelligence, motivation, pitch discrimination, and certain biographic elements. While the predictive accuracy they afford is satisfying, it does not come about in the way one might wish. It had been hoped that foreign language achievement could be predicted on the basis of *intellectual* factors, such as the ability to discriminate sounds, to induce grammatical principles, and so on. Instead, it appears from these studies that the two biggest factors in such achievement are the very general ones of verbal IQ and motivation. This finding says that as far as language study in college is concerned, anyone will do well who is intelligent and wants to learn, regardless of such concerns as having "a good ear," "a good memory," and "good reasoning powers."

Nonetheless, it is of interest to examine the relative importance of the factors which contribute to the prediction of each criterion.

Though the first criterion (the grammar-reading goal) was measured in quite different ways in the two studies, this fact will be ignored for the moment, to permit comparison of the two sets of findings. These are the tests which contributed 1% or more to prediction of the first criterion:

[7] Shrinkage estimated by McNemar's Formula 75. See McNemar (1955, p. 186).

Study I		
3.	Words in Sentences	.047
1.	Spelling Clues	.033
2.	Number Learning	.025
19.	Bilingualism	.022
6.	Reading Aloud I	.021
9.	Reading Aloud II (errors)	.014
17.	Age	.012
15.	Linguistic Analysis II	.010
Study II		
1.	Verbal Comprehension	.193
13.	Interest I	.123
17.	Sex	.022
14.	Interest II	.020
18.	High School Language	.019
10.	Chinese Pitch	.019
2.	Linguistic Analysis I	.017
6.	Rhymes	.010

In Study I, three of the tests (3, 2, 15) are associated with the Reasoning factor. Test 1 is associated with the Verbal factor. Test 6 is associated with Speed of Articulation; Test 9 with Response Set; Test 17 with the Biographic factor.

The factorial findings in Study II differ partly from those of Study I, since different tests are involved. In Study II, Tests 1 and 18 are associated with a Verbal factor; Tests 13 and 14 with an Interest (motivation) factor; Test 17 with a Biographic factor; Test 10 with a Pitch Discrimination factor; Test 2 with a Reasoning factor; Test 6 with a Word Fluency factor. The Verbal, Reasoning, Word Fluency, and Biographic elements were found in both studies. The interesting Speed of Articulation finding of the first study was not confirmed by the second. Interest and Pitch Discrimination are new additions in the second.

These findings permit the tentative conclusion that achievement of the traditional (grammar, reading, writing) objectives of a college French course involve primarily verbal intelligence and motivation on the student's part, but may also involve his ability to reason both analytically and by analogy, and his ability to think up words quickly.

In order to examine the two sets of findings with regard to oral (speaking) achievement, the difference between the two measures of the criterion will again be ignored. There follows a summary of the tests which contribute 1% or more to the prediction of this criterion:

Study I		
1.	Spelling Clues	.058
16.	Verbal Comprehension	.036
19.	Bilingualism	.028
6.	Reading Aloud I	.023
5.	Letter Series	.019
8.	Reading Aloud II	.010
Study II		
18.	High School Language	.048
13.	Interest I	.040
1.	Verbal Comprehension	.021
8.	Reading Aloud II	.018
19.	High School Math-Science	.016
6.	Rhymes	.015
17.	Sex	.013

Represented in the findings of Study I are the Verbal factor (Tests 1 and 16), the Speed of Articulation factor (6 and 8), and the Reasoning factor (5). The factors represented in Study II are Verbal (Tests 1 and 18), Interest in Language (13), Speed of Articulation (8), Reasoning (19), Word Fluency (6), and Biographic (1, 17, 18).

These findings suggest that in order to do well in learning to *speak* French in college the student must possess, above all, verbal intelligence and high motivation, and that in addition he may be helped by his analytic reasoning ability, his ability to articulate words rapidly, and his ability to think them up.

As for the third criterion, listening comprehension, it is predicted by:

Study II		
1.	Verbal Comprehension	.075
13.	Interest I	.032
2.	Linguistic Analysis I	.030
10.	Chinese Pitch	.011
12.	Seashore Timbre	.009

The factors involved here are, again, verbal intelligence and motivation, plus reasoning, pitch discrimination, and timbre discrimination.

It is concluded from the studies thus far completed (*a*) that achievement in a traditional (grammar-reading) language course may be predicted with reasonable validity by a set of tests, each of which taps only one rather precise characteristic of the learner; (*b*) that nonintellectual characteristics, notably motivation, must be included as well as intellectual ones; (*c*) that oral and aural achievement are less subject to satisfactory prediction at the present time, probably due to the lack of adequate criterion tests for achievement in these skills; (*d*) that although better criterion tests will improve prediction somewhat, substantial improvement probably demands the inclusion of entirely new factors as predictors; (*e*) that among such new factors, the personality of the student and the characteristics of the teacher are those which appear most promising and are most in need of research attention.

REFERENCES

CARROLL, J. B. A factor analysis of verbal abilities. *Psychometrika*, 1941, **6**, 279-307.

CARROLL, J. B. A factor analysis of two foreign language aptitude batteries. *J. gen. Psychol.*, 1958, **59**, 3-19.

CARROLL, J. B. The prediction of success in intensive foreign language training. Harvard University Graduate School of Education, Laboratory for Research in Instruction, 1960. (Mimeo)

CARROLL, J. B., & SAPON, S. M. *Modern Language Aptitude Test*. New York: Psychological Corporation, 1955.

GUILFORD, J. P., & CHRISTENSEN, P. R. A factor-analytic study of verbal fluency. *U. Sth. Calif. Psychol. Lab. Rep.*, 1956, No. 17.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187-200.

MCNEMAR, Q. *Psychological statistics*, New York: Wiley, 1955.

PIMSLEUR, P. A French Speaking Proficiency Test. *French Rev.*, 1961, **34**, 470-479.

THURSTONE, I. L., & THURSTONE, T. G. Factorial studies of intelligence. *Psychometr. Monogr.*, 1941, No. 2.

(Received April 7, 1961)

Journal of Educational Psychology
1962, Vol. 53, No. 1, 27-31

TRAINING ADULTS TO JUDGE CHILDREN'S SOCIAL ACCEPTANCE

HELEN R. MARSHALL
Drexel Institute of Technology

To determine whether specific training improves the accuracy of adult judgments of children's social acceptance, 290 college students enrolled in a course requiring observation of a preschool group were divided into a control group, a group trained in factual knowledge, and a group exposed to a standard as well as to factual knowledge. Accuracy of judgment was tested 6 weeks after training. Results suggested that reduction as well as improvement in accuracy may result from training. Accuracy decreased when information as to earlier sociometric choices was provided, and increased after factual knowledge training. IQ was related to accuracy of judgment after training limited to factual knowledge, while personality characteristics influenced judgments only when no training was given.

Investigations have shown that most teachers are grossly inaccurate judges of the social acceptance of children in their classes, but have not revealed either bases for easy identification of the few accurate teachers, or how teacher judgments could be improved.

The hypothesis that specific training improves accuracy seems as applicable to judgments of social acceptance as to other learning. Training about factors influencing children's acceptance by peers has not been included in the preparation of most teachers. A first purpose of this investigation was to test this application.

The study was planned to test the specific hypothesis that more adults will reach a minimum accuracy level (agreement with children's choices beyond the .05 level of significance) if they are given training about the social behavior relating to social acceptance by preschool peers, and are exposed to earlier sociometric choices of the children to be judged. The factual demonstration planned for the second phase of this hypothesis derived from a suggestion by Gronlund (1950) that teachers trained in use of sociometric techniques might judge more accurately.

A second purpose of the investigation was to study the effect of training conditions on relations between accuracy of judgments and both IQ and certain personality characteristics of the judges. This purpose required maximizing the range of scores in training groups.

METHOD

Subjects

Initiation of this investigation depended upon the validity of an assumption that college students enrolled in a course requiring weekly observation of a nursery school group might be as accurate in their judgments as the teachers studied in past research. If half had better than chance agreement with children's sociometric choices, they would be comparable to teachers. This conjecture was supported (Marshall, 1958) when tested on the untrained control group of this study.

The 290 subjects were about two-thirds of the enrollment in a child development course of the University of Kentucky[1] in seven of the eight semesters from fall 1956, through spring 1960. Students were discarded as subjects if they were absent at test or training sessions, had not taken the university entrance examination, or had a score of seven or more on the Lie scale admin-

[1] The data of this report were collected and partly analyzed while the author was a faculty member of the University of Kentucky.

istered with the Manifest Anxiety scale (Taylor, 1953). Difficulties with photographic arrangements needed for the children's sociometric test prevented training students in the fall 1957 semester.

Students comprising a year's enrollment in the course had a wider range of scores than was found in any semester, section, or projected subdivision. For this reason, students were assigned to training groups on the basis of date of enrollment in the course, as is shown in Table 1. A change in entrance examination scores was the only difference between training groups in tested personal characteristics. Mean percentile rank was 35 in 1956-57 on the Kentucky Classification Battery, but it was 50 in 1959-60. The lower limit of the range increased from 1 to 25. The change took place during the 2 years of comprehensive training. The upper limit was 98 or 99 in all groups.

A change in nursery school enrollment, shown also in Table 1, meant that students split their 2 weekly hours of observation between preschool groups during the last 2 years, and had less time to observe each child.

Training

The two trained groups were given the same introduction and description of nursery school behavior relating to preschool social acceptance. Training was conducted by the investigator in one lecture period during the fifth or sixth week of the semester. Students were informed that they were subjects in a study about the effect of training on judgments. Research findings about the accuracy of teacher judgments were summarized. Knowledge of the criterion for their accuracy was given by a description of the method of obtaining children's choices, the picture sociometric test (McCandless & Marshall, 1957).

Description of social behavior relating to choice as a playmate by preschool peers required 30 minutes. Content was based on research done by the investigator, and emphasized the following relations.

1. Sociometric choices depend more directly on the extent of two types of observable social participation than on actual playmates. A child is more often chosen by other children in the group when the child spends more time talking and acting out the roles of dramatic play with them, and the child interacts with a larger number of children in a given time. These two behaviors tend to increase with age, but age per se does not influence choices. If a girl and a boy appear equal in these behaviors, the girl is more often chosen as a playmate by peers.

2. Acceptance in a preschool group is not influenced by (*a*) the extent of aggression and hostility shown to peers, and (*b*) how often the child talks to others as himself, rather than as an imaginary thing or person. An exception to *a* is that peer choices of boys increase as boys more often express hostility in acting out dramatic play roles.

3. A child who is more dependent on the teacher during play than others of the same age will be less often chosen as a playmate by peers.

4. Resemblance between sociometric test choices of preschool children decreases as the time between tests increases from 1-6 weeks. Choices on the first test are not good predictors of choices made 6 weeks later.

Behavior examples were drawn from research observations and from observation of the current nursery school group.

No other training was given to the *limited training group*. The remaining time was spent discussing development of social behavior, a topic considered at other class meetings for the control and comprehensive training groups.

The *comprehensive training group* received the *additional* training of a demonstration of sociometric choices of the preschool children they were observing and later would judge. Children had been tested during the week preceding training. Preferred playmates of each child were listed on the blackboard before the session, and remained there during the 50-minute period. An oral summary of the choices was presented after the introduction and before the description of social behavior. The fact that these choices were a useless basis for judgment at the end of the semester was strongly emphasized.

Testing

All subjects judged the social acceptance of the children twice. The first test period occurred in the week preceding training, and the second occurred 6 weeks after training, or after 11 or 12 weeks of observation.

Students were asked to list the three children they thought each child would select as best friends on the child's sociometric test. Their tests were compared with the sociometric tests given to the children during the same week.

The probability of the agreement between

each student's judgment and each child's choices, irrespective of order of choice, was determined for the specific preschool group, and converted into the appropriate chi square. If the sum of the chi squares was large enough for significance at the .05 level, the student was classified as judging at an accuracy level beyond that of chance.

Tests of personality characteristics administered at either or both test periods to all groups are listed in Table 2.

RESULTS

The results shown in Table 1 indicate that accuracy of adult judgments of children's social acceptance can be both improved and reduced by specific training.

Training limited to a description of social behavior relating to acceptance by peers was followed by attainment of beyond chance accuracy by all students. This was twice the proportion judging at this level of accuracy in the untrained control group.

When this training was supplemented by exposure to past sociometric choices of the children to be judged, the proportion of students judging at the .05 level of accuracy decreased to half that of students given no training, and to one-fourth that of students given the limited training.

According to this evidence, the application of the specific training hypothesis to adult judgments of social acceptance depends upon the content of specific training. Additionally, the hypothesis needs to be modified to include the possibility of effects in either direction.

Insight into the reasons for these differences is provided by the correlations between percentage accuracy of student judgments and personal characteristics scores shown in Table 2. Entrance examination scores correlated with percentage accuracy only within the limited training group. The apparent relation in the case of the control group was indicated by partial r's to be due to the influence of manifest anxiety scores (described further on p. 210, Marshall, 1958). A relationship between IQ and

TABLE 1

COMPARISON OF TRAINING GROUPS ON DETERMINATIONS OF ACCURACY OF JUDGMENT OF PRESCHOOL CHILDREN'S SOCIAL ACCEPTANCE

Determination of accuracy	No training 1956-57 ($N = 131$)	Comprehensive training			Limited training 1959-60 ($N = 47$)
		Total ($N = 112$)	Spring 1958 ($N = 50$)	1958-59 ($N = 62$)	
Percent of subjects judging at .05 level of accuracy**	57	23	0	42	100
Percent of subjects judging at .01 level of accuracy**	23	8	0	15	89
Mean percentage agreement of student and child choices	47	35*	32	37*	37
Number of groups judged	1		1	2	2
Number of children judged	13		12	18 and 20	29 and 33
In younger group				12 and 13	15 and 17
In older group				6 and 7	14 and 16

* Percentage for small N of older group adjusted to that of a group of 12 children.

** All differences between three training groups were significant at the .001 level by χ^2 test.

TABLE 2

PRODUCT-MOMENT CORRELATIONS BETWEEN PERSONAL CHARACTERISTICS SCORES AND PERCENTAGE AGREEMENT AT THE END OF THE SEMESTER OF STUDENT AND CHILDREN'S CHOICES IN THREE TRAINING GROUPS

Personal characteristics scales	No training ($N = 131$)	Comprehensive training			Limited training ($N = 47$)
		Total ($N = 112$)	Spring 1958 ($N = 50$)	1958-59 ($N = 62$)	
College entrance examination	.24*[b]	−.04[a]	−.19	.09	.43*[b]
Manifest Anxiety scale	−.48*[a]	−.02[b]	.07	−.09	.09[b]
Harsh Punitive Control scale of Parent Attitude Research Instrument[c]					
Early in semester		−.13	−.11	−.15	−.09
At end of semester	−.32*	−.15	−.17	−.14	−.06

[ab] Difference between r^a and r^b in the separate rows of the table significant at .05 level.

[c] Schaefer and Bell (1955).

* Significant at .01 level.

accuracy of judgments did not exist for judgments any group made prior to training. IQ influenced learning to judge social acceptance, then, only after training limited to knowledge of social behavior relating to social acceptance.

Both types of training nullified the influence of other personality characteristics on accuracy of judgments at the end of the semester, according to correlations shown in Table 2. Negative correlations between personality test scores and percentage accuracy at the end of the semester were found in the group given no training, and in all three groups before training.

Two possible explanations of the reduction in accuracy obtained after the comprehensive training were explored through additional analyses.

The possibility of excessive reliance on the sociometric choices presented during training was studied by determining the percentage agreement of students' end of semester judgments and children's early in semester choices. In the spring 1958 group, agreement was 44%, and was higher (.001 level by t test) than that shown in Table 1. In the next year of comprehensive training, agreement was 36%, and did not differ from that shown in Table 1. Fewer than 5% of the students in any group made judgments early in the semester that agreed beyond the .05 level with children's early in semester choices. These data support the possibility that students relied unduly on children's choices presented in training.

This possibility also is supported by the evidence in Table 1 that students given comprehensive training were less accurate in end of semester judgments in the year when they judged fewer children than in the year when they judged more children.

The possibility that students in the comprehensive training group might not have used observation cues in forming judgments was studied through use of data collected for other purposes. The most frequent playmates of each child during the 4 weeks preceding the final sociometric tests were derived from ob-

servation records for the preschool groups judged by the control group and the spring 1958 comprehensive training group. Agreement at the .05 level of accuracy was attained by 38% of the comprehensive training group, and by 80% of the control group. Clearly, the comprehensive training students relied less on observed behavior than students given no training.

DISCUSSION

Learning a social behavior (judgment of social acceptance of children) was shown to relate to IQ in one, but not both training conditions. The conditions required for influence of IQ appeared to be presentation to all subjects of a factual basis for judgment, and the absence of exposure to examples or a standard which may, in effect, provide misinformation. A higher IQ did not facilitate learning prior to training, and it did not prevent the reduction in accuracy found to accompany exposure to a standard, i.e., knowledge as to earlier sociometric choices made by the pupils under observation.

Reduced accuracy of judgment was a consequence of the type of training that was used to "make situations real" to students. It is widely believed that students learn more when factual training is supplemented by examples of real situations. Under the conditions employed in this investigation, this supplement sufficed to eliminate both the facilitating effects of factual knowledge, and cues from prior experience, which presumably were used by students in the control group.

Analyses supported two conjectures about concomitants of the reduced accuracy: undue reliance on the examples presented in training, and decreased use of observed behavior as a basis for judgment. Perhaps adults in a training situation hesitate to use their own judgment after another adult has pronounced judgment, or given a standard.

It is concluded that training can improve the accuracy of teachers' judgments of children's social acceptance, but that further study is desirable to determine the types of training that will be most advantageous.

REFERENCES

GRONLUND, N. E. The accuracy of teachers' judgments concerning the sociometric status of sixth grade pupils. *Sociometry*, 1950, **13**, 197-225, 329-362.

MCCANDLESS, B. R., & MARSHALL, HELEN R. A picture sociometric technique for preschool children and its relation to teacher judgments of friendship. *Child Develpm.*, 1957, **28**, 139-147.

MARSHALL, HELEN R. Personality characteristics of college students and accuracy of their judgments of children's social acceptance. *J. home Econ.*, 1958, **50**, 207-212.

SCHAEFER, E. S., & BELL, R. Q. Parental Attitude Research Instrument (PARI): Normative data. Unpublished manuscript, Library, National Institutes of Health, Bethesda, Maryland, 1955.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.

(Received May 22, 1961)

Journal of Educational Psychology
1962, Vol. 53, No. 1, 32-37

THE EFFECTS OF DISCRIMINATION PRETRAINING ON TEXTUAL BEHAVIOR[1]

CAROLYN K. STAATS, ARTHUR W. STAATS, AND RICHARD E. SCHUTZ
Arizona State University

The study compared 3 groups (12 each) of kindergarten children in learning textual responses. Group 1 had discrimination pretraining with the same words as those in the criterion task, Group 2 with the letters making up these words, and Group 3 no discrimination training. The response analyzed was the number of correct anticipations in 10 trials of a 4-word task. Analysis of variance showed no difference between groups, but a Duncan new multiple range test indicated a significant difference between Group 1 and Groups 2 and 3, but not between the latter. Thus there was positive transfer to a later task involving the same stimuli, but this did not seem due to acquisition of an observing response to the components of the stimulus.

Muehl (1960) has investigated the effects of different types of visual discrimination pretraining on the performance of kindergarten children in learning to read a vocabulary list. He found that children who received pretraining in matching the same words that appeared in a later vocabulary list learned to "read" the words significantly better than children who received similar pretraining with different words or with geometric forms.

Muehl suggests that it was the response to single letters or details of letters that appeared to account for his finding. He hypothesizes that "positive transfer to the learning task . . . was due to the acquisition, during pretraining, of a set of observing responses to distinctive letters in the stimulus words."

The present study extends the stimulus pretraining research by comparing the effects of two different types of pretraining experiences. Prior to learning the same vocabulary list, kindergarten children were exposed to one of three conditions: (*a*) discrimination pretraining with the same words as those appearing in the vocabulary list, (*b*) discrimination pretraining with the letters making up the words in the vocabulary list, and (*c*) no discrimination training.

Based upon the hypothesis that if differences are established among similar stimulus elements, then learning will be facilitated in a subsequent task using these same stimuli, it would be predicted that no significant difference in learning the vocabulary list would be found between the group given visual discrimination pretraining on the same words and the group receiving visual discrimination pretraining on the letters in the words. Both these groups would be expected to learn the vocabulary list faster than the group given no discrimi-

[1] The research reported in this paper was supported by the United States Office of Education, Cooperative Research Branch, under Contract No. 1048. A summary of this article was presented at the national meetings of the American Psychological Association in New York City, September 1961. The authors wish to thank John Mabry and Donald V. Huard for help in collecting the data and Irwin Miller for aid in interpreting the statistical analysis. The authors also wish to express gratitude for the assistance of Roy Doyle and Carol Steers of the Arizona State University Campus Laboratory School, Lois Kline of the District Office of Tempe Elementary Schools, Frank Huston and Jacqueline Miller of Broadmor Elementary School, Augusta Carminati and Patricia Huddle of Wayne Ritter Elementary School, and Robert Curry of McKemey Junior High School in Tempe, and Lorenzo Rich of West Mesa Junior High School in Mesa.

nation pretraining. On the other hand, if the requisite condition for facilitation of learning is discrimination pretraining on the same stimuli, it would be predicted that the group given training on the same words would learn the vocabulary list faster than either of the other two groups.

Another line of theory and research also contributes to the present study. Skinner (1957) introduced the term "textual response" to refer to an operant vocal response under the control of nonauditory verbal stimulus, i.e., printed symbols. It is suggested that this term should replace the phrase "reading aloud" because the complicated connotations of the verb "to read" are often misleading. Questions of "comprehension" or "understanding" or "meaning" frequently become spuriously intertwined with research aimed at understanding the variables involved in learning to make the appropriate vocal response to the printed verbal stimulus. Although a full analysis of reading must include the principles of classical conditioning and the interaction of classical and operant principles in the development of word meaning (Staats, 1961; Staats & Staats, 1957), the use of the term texting concerns only the operant conditioning of a vocal response under the control of a visual verbal stimulus. The present study utilized these principles in two ways: discrimination pretraining was programed (the discrimination progressing in difficulty through the training with little opportunity for failures); and reinforcement was made contingent upon the emission of correct vocal operants in both discrimination and texting training.

METHOD

The procedure was composed of three parts: familiarization training which was uniform for all subjects, discrimination pretraining which differed for each group, and criterion texting which was the same for all subjects.

Subjects

The subjects were 36 children from two kindergartens in Tempe, Arizona. Shortly prior to the experiment the Columbia Mental Maturity Scale was administered to all subjects. The children were then matched on MA and assigned to one of the three pretraining groups. There were 12 children in each group. Table 1 shows the mean ages and MAs in each group. In all phases subjects worked individually with one of two adult male experimenters.

Familiarization Training

The following procedure was used to acquaint the subjects with the token reinforcement system and to train them in the responses required for the discrimination task. To establish the tokens as reinforcers, the subject was given four or five token plastic discs and allowed to choose and buy a toy (plastic trinket). This was done twice for each subject. The training for the discrimination task involved the following: Subjects were shown 10 cards consecutively, each containing three stimulus components (geometric figures, drawings of common objects, or combinations of letters printed upside down), two of which were the same and one different. The first five cards were simply 5 × 8 replications of the first five items of the Columbia Mental Maturity Scale with which the children were already familiar because of prior testing. The subject was instructed. "Point to the one that doesn't belong." If the subject responded correctly,

TABLE 1

MEANS AND *SD*s OF MENTAL AND CHRONOLOGICAL AGE IN MONTHS

	Group I		Group II		Group III	
	M	*SD*	*M*	*SD*	*M*	*SD*
Mental age	65.9	7.07	65.8	5.29	65.4	4.90
Chronological age	67.0	2.64	64.7	1.73	66.9	3.32

he was given a token, while if the response was incorrect or if the subject failed to respond in 5 seconds, the experimenter withheld the token reinforcer and presented the next card. At the end of the series, the tokens were once again exchanged for toys.

Discrimination Pretraining

The two different sets of stimuli used in the pretraining phase each consisted of 48 additional cards of the same type of task as that used in the familiarization training. Reinforcement in the form of tokens, later exchanged for toys, was again made contingent upon the correct response.

For Group I the discriminative stimuli consisted of words printed in lower-case type which were later used in the texting task. Twenty-four cards were made up of combinations of the four-letter words: *door, bell, fish, kite.* One of these four appeared twice, another appeared once on the card, and the subject's task was to "point to the one that doesn't belong." Every word was paired with each of the other three in this series from one to three times. The remaining 24 cards each contained one of the same four-letter words paired once with other four-letter words not included in the texting task. Of these 24, eight discriminations involved words with all four letters differing, eight with three letter differences, and eight with two letter differences. The combined 48 cards were then arranged progressively as to difficulty of the discrimination, e.g., *fish* and *door* appeared on a card early in the program, while *fish* and *film* were paired on Card 47. The position of the correct alternative on the card was varied across the series of 48 cards to prevent a position preference.

For Group II the 12 lower case letters *b, d, e, f, h, i, k, l, o, r, s,* and *t,* which comprise the words *door, bell, fish,* and *kite,* were presented individually rather than the words themselves. Along with each single letter another letter, randomly selected from the same list of 12 letters, appeared twice. Each of the 12 letters in the list appeared as a correct alternative on four occasions, except for *e* which occurred five times, and *k* which was presented three times. These cards were then arranged progressively as to difficulty of discrimination, e.g., Card 1 was *k k o,* while Card 42 was *e o e.* For the series of 48 cards the location of a correct letter was varied so as to avoid a position preference. The subject's task was again to "point to the one that doesn't belong."

In both these groups, tokens were exchanged for toys subsequent to Trials 8, 16, 32, and 48.

Group III received no discrimination training other than that implicit in the familiarization training.

Texting Task

Immediately following the pretraining discrimination task, the same texting task was administered to all subjects. This criterion texting consisted of 12 presentations of each of the four words presented to Group I in the discrimination training (*door, bell, fish, kite*). These words were organized into six blocks, each block consisting of two presentations of each of the four words; the order of the words within each block was randomized. Each block was then divided into two trials of four words each. This yielded 12 trials in all.

Responses on the first trial (Cards 1-4) were pointing to the word as the experimenter said the word, and for the second trial (Cards 5-8) the subject was to point to the word and repeat the word aloud (an "echoic" response). Since these two trials permitted no variability among subjects, the first block (Cards 1-8) is not included in the analysis.

The final 10 trials (or five blocks) constituted the criterion task. The subject was instructed by the experimenter to "say the word before I say it," constituting a paired-associate method of presentation. The experimenter presented the stimulus card, and then waited approximately 2 seconds for the subject to say the word. Correct responses were reinforced as before, and incorrect responses were prompted but not reinforced. Tokens were exchanged for trinkets at the end of Trials 7 and 12. The entire experimental procedure was conducted in a small room close to the subject's own classroom.

RESULTS

Learning curves depicting the number of correct anticipations occurring in the texting task for each of the pretraining groups are presented in Figure 1. A Bartlett's test (Lindquist,

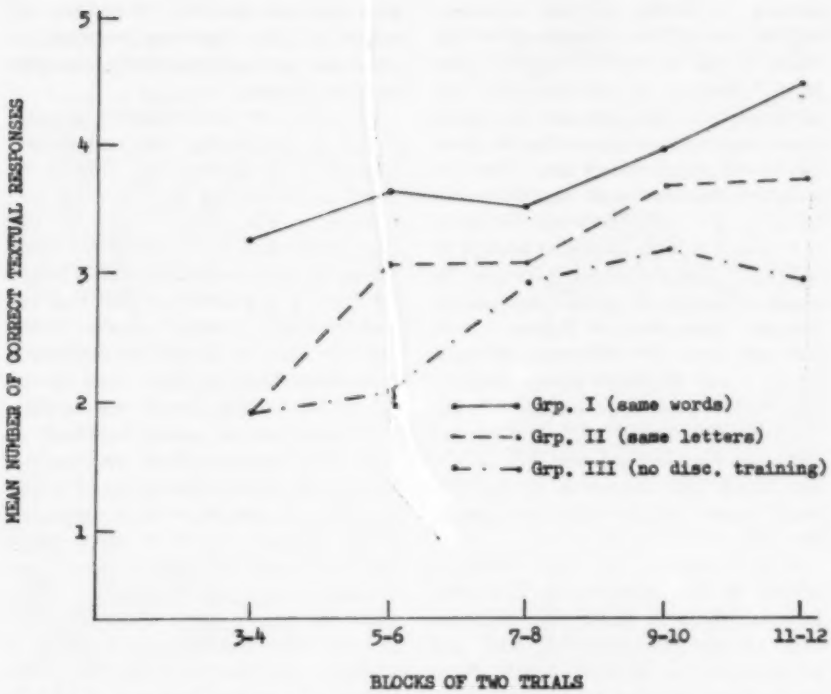


FIG 1. The effect of type of discrimination pretraining on correct textual responses in learning the criterion task. (Eight responses were possible in each block of two trials.)

1953) showed that the variance in the number of correct responses in the three groups was homogeneous. A summary of the analysis of variance (Lindquist, 1953, Type I) for these data is given in Table 2. The F of 6.77 ($p < .01$) for Trials shows that significant learning occurred in the course of the texting training over the three groups. The F of 1.42 for Groups indicates that the type of discrimination pretraining did not significantly affect texting performance.

Edwards (1960, p. 136) points out, however, that "in making multiple comparisons among treatment means, it is not necessary that the treatment mean square of the analysis of variance be significant." He suggests that the choice of multiple comparison method should be guided by questions of experimental

TABLE 2

ANALYSIS OF VARIANCE OF CORRECT TEXTUAL RESPONSES

Source	*df*	*MS*	*F*
Between subjects	35		
Groups (B)	2	21.06	1.42*
Error (b)	33	14.82	
Within subjects	144		
Trials (A)	4	11.71	6.77**
A × B	8	1.17	.64
Error (w)	132	1.73	

*$p > .20$.
**$p < .01$.

interest. According to this rationale, and because of the discrepancy of the results of the analysis of variance with Muehl's findings, it was considered appropriate to further compare the means of the pretraining groups with the Duncan new multiple range test. This test indicated that there were significant differences ($p < .05$) between the mean of Group I and the means of Groups II and III. The difference between the means of Groups II and III was not significant. Inspection of Figure 1 also indicates that the difference between Group I and the other groups was evident by the first block included in the analysis. The mean number of correct responses for Group I was 3.25 in the first block, but the mean of the first block in each of the other two groups was 1.83.

Correlations were also obtained among the MA scores on the Columbia Mental Maturity Scale, the performance on the discrimination task, and performance on the texting task. None of the three correlation coefficients was significant.

DISCUSSION

The results tend to corroborate the findings of Muehl and other investigators (Kurtz, 1955; Spiker, 1956) that stimulus discrimination pretraining results in positive transfer to a later task when the same stimuli are presented in that task. The interpretation that these results are due to acquisition of an observing response to the distinctive letters making up the words was not supported since no significant difference was found between the group receiving pretraining on letters making up the words and the group receiving no pretraining at all.

It would seem, then, that the word functioned as a unique stimulus in the discrimination training. Whatever the nature of the observing response, it was made to the entire word, a complex unitary stimulus.

However, the results cannot be interpreted as indicating that alphabetical pretraining in general has little or no effect in enhancing later texting performance. The subjects used in this experiment had not learned the letter names. Subjects were deliberately selected from a population which had not had "alphabet training" in order to isolate the effect of simple discrimination pretraining, and to insure that the response of naming would not mediate the acquisition of textual responses. It may be, however, that texting pretraining on alphabetical units of words, including the development of some phonetic responses to these units would prove the most efficacious type of pretraining in the applied situation. Thus, while the results question the relevance of pure discrimination pretraining on alphabet components of words in facilitating later texting behavior, it may be that texting training on these units would prove valuable for transfer to the whole word. Further research relating texting of letter units to the later texting of entire words thus would appear to be fruitful.

Another reason for questioning the efficacy of pure discrimination pretraining for the later texting behavior is the wide variability in the texting performance of subjects in the same group shown in Table 2. Since these variances were homogeneous, the size of the *MS* for between subjects in the same group may be interpreted as indicating the large differences among individuals were maintained in spite of constant experimental treatment. Thus, even though the results of Group I were, overall, superior to the other groups, large differences in the criterion texting

are still attributable to other unknown variables. Even the use of programed discrimination training did not suffice to bring the children to a common starting level.

Finally, the lack of a significant correlation between discrimination and texting behavior indicates that better discrimination performers were not better learners of texting. This indicates that the same behaviors were not involved in the two tasks and that the acquisition of texting does not depend solely on the previous acquisition of observing responses.

The lack of significant relationship between the pretest performance, i.e., the Columbia Mental Maturity Scale and the other two variables, also suggests that different variables are involved in performance of these tasks. The use of these types of tests for predicting early reading behavior thus seems questionable.

REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1960.

KURTZ, K. H. Discrimination of complex stimuli: The relationship of training and test stimuli in the transfer of discrimination. *J. exp. Psychol.*, 1955, **50**, 283-292.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MUEHL, S. The effects of visual discrimination pretraining on learning to read a vocabulary list in kindergarten children. *J. educ. Psychol.*, 1960, **51**, 217-221.

SKINNER, B. F. *Verbal behavior.* New York: Appleton-Century-Crofts, 1957.

SPIKER, C. C. Experiments with children on the hypothesis of acquired distinctiveness and equivalence of cues. *Child Develpm.*, 1956, **27**, 253-263.

STAATS, A. W. Verbal habit-families, concepts, and the operant conditioning of word classes. *Psychol. Rev.*, 1961, **68**, 190-204.

STAATS, C. K., & STAATS, A. W. Meaning established by classical conditioning. *J. exp. Psychol.*, 1957, **54**, 74-80.

(Received May 25, 1961)

Journal of Educational Psychology
1962, Vol. 53, No. 1, 38-42

SOME CHARACTERISTICS RELATED TO SOCIAL ISOLATION OF SECOND GRADE CHILDREN[1]

HARRIET H. COMMOSS[2]
Teachers' College, Columbia University

Some 20 children scoring in the highest quarter of their classes on a sociometric test were compared with 20 children in the lowest quarter with respect to the following characteristics: learning of sex role, certainty in interpersonal relationships, ability to communicate verbally, and eye-hand coordination. Positive relationships between social status and the stated characteristics were hypothesized. Each child copied a short sentence and was presented with the Driscoll Playkit and asked to complete 10 stories. Analysis of the data failed to support the hypothesis of a positive relationship between social acceptance and learning of own sex role but reliably supported the hypotheses related to the other characteristics.

Harry Stack Sullivan (1953), in developing his theory, laid emphasis on the importance of the early school years and the effect of social isolation at that period. He considered this period to be the actual time for becoming social and consequently believed that anything which interferes with peer relationships then may be doubly handicapping, first in terms of the original faulty learning which results in a greater or lesser degree of social inadequacy, and secondly through the resultant lack of opportunity for corrective or new learning.

He believed that certain areas of inadequacy are usually associated with social isolation. Four of these areas were selected for study, and the following hypotheses were formulated:

Hypothesis 1: Socially isolated children make slower progress than accepted children in personifying themselves as members of their own sex.

Hypothesis 2: Socially isolated children are more uncertain as to reactions of others toward them and the outcomes of situations than are accepted children.

Hypothesis 3: Socially isolated children tend to be more restricted in freedom of communication than accepted children.

Hypothesis 4: Socially isolated children progress more slowly in the development of eye-hand coordination than accepted children.

METHOD

Subjects

In order to test the postulated relationships, two groups of second grade children were selected. They were chosen by means of a sociometric questionnaire from the members of 10 classes in five schools in a small New England city. Each child was asked to choose three others for each of three purposes: to sit beside, to play with at recess, to invite home after school. The responses to the questionnaire were weighted in line with the procedure described by Northway (1940). First choices received a weight of five, second choices of three, and third choices of two. The choices received by each child were then added together, and from the distribution of scores so obtained, two groups of children were selected for further consideration. Those in the top quarter of their own class distributions were considered to represent children highly acceptable to their peers and are referred to henceforth as accepted children. Position in the lowest quarter was held to indicate some degree of social isolation, and children whose scores fell in this portion of the distribution are termed isolated children.

[1] This article was taken from a report with the same title submitted in partial fulfillment of the requirements for the degree of Doctor of Education, September 1, 1960.

[2] At present: School Psychologist, Great Neck, New York.

In the final selection of 20 children from each of the above groups, an effort was made to match the groups as nearly as possible in socioeconomic status and in intelligence. No child was included whose family received a rating lower than five (semiskilled worker) on the Edwards (1938) scale of socioeconomic grouping, and the IQs of all subjects as measured by the Pintner-Cunningham test, were between 90 and 120. The average IQ of the accepted children was 104.0, while that of the isolated children was 105.9 ($r = -.18$).

Each child was presented with the Driscoll Playkit and 10 incomplete stories were read to him. The Playkit is an instrument developed to facilitate the use of doll play in studying children. It consists of a small wooden suitcase containing furniture and partitions needed to set up a six-room doll house. A doll family of five is a standard part of the equipment. The kit is described in detail by Walsh (1956) together with directions for presentation of material and recording of data.

The stories to which the children were asked to respond briefly described some situation which was likely to be within their usual experience such as the following: "The child has just done something which the parents said not to do. What do you suppose happens next?" They were invited to answer both in words and through play with the dolls. The responses were recorded on tape and transcribed verbatim.

A sample of each subject's manuscript writing was obtained by asking him to copy the following sentence: "Betty is a little girl."

Analysis of Data

The stories were analyzed in the following ways: proportion of responses referring to the children dolls which attributed action, thinking, or feeling to the child doll of the same sex as the child responding; ability or inability to respond with definite consequences to the story situations presented (scores for each story of 0, 1, or 2 determined with the assistance of judges); and number of words spoken per minute. The writing samples were rated as "satisfactory" or "poor" by seven primary grade teachers.

RESULTS

It was found that no significant difference existed in the relative frequency with which children in the experimental and control groups referred to the children dolls of their own sexes. Although the results shown in Table 1 are in the direction hypothesized, they failed to reach the .05 level of significance when evaluated by the median test for two samples.

The socially isolated children were distinctly less willing than the accepted to commit themselves as to the outcomes of situations or the reactions of characters in their stories. The mean scores

TABLE 1

PROPORTION OF RESPONSES REFERRING TO CHILD DOLL OF SAME SEX AS CHILD RESPONDING

Social status	Number of scores	
	Above median	Below median
Accepted		
Boys	6	4
Girls	5	5
Both sexes	11	9
Isolated		
Boys	7	3
Girls	2	8
Both sexes	9	11

TABLE 2

CHILDREN'S SCORES AS DETERMINED BY CONSEQUENCES DESCRIBED IN STORIES

Social status	Mean scores
Accepted	
Boys	13.6
Girls	14.1
Both sexes	13.85*
Isolated	
Boys	11.6
Girls	11.6
Both sexes	11.6*

* The difference between these means is significant at the .01 level.

on these variables for both groups are shown in Table 2. The difference between means is significant ($t = 2.68$, $df = 38$, $p < .01$).

The accepted children proved to be more fluent than the isolated children in responding to the stimuli presented. When the data shown in Table 3 were evaluated by the t test, the difference between groups was found to be significant ($t = 1.85$, $df = 38$, $p < .05$).

TABLE 3

RATE OF TALKING IN RESPONSE TO INCOMPLETE STORIES

Social status	Mean number of words per minute
Accepted	
Boys	71.6
Girls	88.4
Both sexes	80.0*
Isolated	
Boys	54.8
Girls	70.7
Both sexes	62.8*

*The difference between these means is significant at the .05 level.

TABLE 4

RATINGS OF HANDWRITING AS DETERMINED BY AGREEMENT AMONG FIVE OUT OF SEVEN JUDGES

Social status	Number of samples rated		
	Satisfactory	Poor	No agreement
Accepted			
Boys	7	0	3
Girls	4	5	1
Both sexes	11*	5*	4*
Isolated			
Boys	1	8	1
Girls	3	5	2
Both sexes	4*	13*	3*

*Significant at .05 level.

The data shown in Table 4 lend support to the hypothesized relationship between sociometric rating and eye-hand coordination as shown in handwriting ($\chi^2 = 6.96$, $df = 2$, $p < .05$).

DISCUSSION

Since the data of this study failed to support the hypothesis that socially isolated children make slower progress in learning their appropriate sex roles than children who are accepted, a question arises as to the adequacy of the experimental measure. Inspection of Table 1, however, shows that the results, insofar as difference between the sexes is concerned, are in line with results of other studies such as those of Brown (1957) and Gray (1957). The boys as a total group gave evidence of a higher degree of sex identification than the girls ($\chi^2 = 3.6$, $df = 1$, $\chi^2_{.95} = 3.8$).

As a further check, the sociometric choices were statistically evaluated. Here, as Table 5 shows, there appeared not only a difference between sexes but also a significant difference between the two groups of boys ($\chi^2 = 8.1$, $df = 1$,

TABLE 5

SOCIOMETRIC CHOICES IN RELATION TO SEX CHOSEN

Social status	Number of children choosing	
	Opposite sex	Own sex only
Accepted		
Boys	0*	10*
Girls	5	5
Both sexes	5	15
Isolated		
Boys	6*	4*
Girls	8	2
Both sexes	14	6

*Significant at .005 level.

$p < .005$). This indicates that the hypothesis may have some validity as far as boys are concerned and suggests the desirability of further experimentation. There is no support, however, for the hypothesis that accepted girls progress more rapidly than isolated girls in the learning of sex role.

The confirmation of both the second and third hypotheses, relating to uncertainty as to other's reactions and freedom to communicate, lends credence to Sullivan's belief that a vicious circle may be created. The anxiety engendered by mistakes due to misconceptions may lead to failure to clarify concepts through open discussion and hence to further misconceptions. Many others including Murphy (1947) and Dollard and Miller (1950) have stressed the damage to personality which may result from such a situation.

In the present study, the hypotheses discussed in the preceding paragraph were tested only in relation to communication with an adult. In view of the fact that much learning at all ages is held to result from interaction with age peers, an important question arises as to whether the findings would have been the same had verbal communication with other children been tested. Studies by Bonney (1955) and Hahn (1948) suggest that this would have been the case.

The confirmation of the hypothesized relationship between eye-hand coordination and social status leads to a more general question as to whether a similar relationship exists at the age studied in respect to motor development in general. Other studies (Hunt & Solomon, 1942; Tuddenham, 1951) suggest that this may well be the case.

An additional exploration was undertaken as a part of this research. The cooperating teachers were asked to make ratings of general coordination on all children remaining in the accepted or isolated groups after those who were ineligible because of intelligence or socioeconomic status had been eliminated. The ratings were on a five-point scale and were based on the teachers' impressions. The results shown in Table 6 makes it evident that a significantly greater proportion of accepted children were considered well coordinated by their teachers ($\chi^2 = 7.59$, $df = 2$, $p < .025$).

CONCLUSIONS

The results of this study lend support to Sullivan's theories concerning characteristics associated with social isolation in the early school years. While the hypothesis of a positive relationship between social status and the learning of sex role was not supported, a question is raised as to the possibility that such a relationship may exist in the case of boys.

The hypotheses concerning the uncertainty of isolated children in regard to the reactions of others and their inhibition in verbal communication were

TABLE 6

RATINGS OF MOTOR COORDINATION AS DETERMINED BY TEACHER OBSERVATION

Social status	N	Number of children rated		
		Above average	Average	Below average
Accepted				
Boys	22	10	6	6
Girls	11	7	3	1
Both sexes	33	17*	9*	7*
Isolated				
Boys	17	4	11	2
Girls	12	2	6	4
Both sexes	29	6*	17*	6*

*Significant at .025 level.

confirmed. It is felt that they acquire special importance when considered in the light of Sullivan's belief that under these circumstances a vicious circle may be created. Due to fear of error and ridicule, the child may fail to expose himself to corrective or new learning with unfortunate results.

A positive relationship between social status and eye-hand coordination was found in the groups studied. Additional evidence and other studies suggest that the same relationship may exist in respect to motor development in general.

REFERENCES

BONNEY, MERL E. Social behavior differences between second grade children of high and low sociometric status. *J. educ. Res.*, 1955, **48**, 481-495.

BROWN, D. G. Masculinity-femininity development in children. *J. consult. Psychol.*, 1957, **21**, 197-202.

DOLLARD, S., & MILLER, N. F. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

EDWARDS, A. E. *A social-economic grouping of the gainful workers of the United States.* Washington, D. C.: United States Government Printing Office, 1938.

GRAY, SUSAN W. Masculinity-femininity in relation to anxiety and social acceptance. *Child Develpm.*, 1957, **28**, 203-214.

HAHN, ELISE. Analyses of the content and form of the speech of first grade children. *Quart. J. Speech*, 1948, **34**, 361-366.

HUNT, J. MCV., & SOLOMON, R. L. The stability and some correlates of group-status in a summer-camp group of young boys. *Amer. J. Psychol.*, 1942, **55**, 33-45.

MURPHY, G. *Personality.* New York: Harper, 1947.

NORTHWAY, MARY L. A method for depicting social relationships obtained by sociometric testing. *Sociometry*, 1940, **3**, 144-150.

SULLIVAN, H. S. *The interpersonal theory of psychiatry.* New York: Norton, 1953.

TUDDENHAM, R. D. Studies in reputation: III. Correlates of popularity among elementary school children. *J. educ. Psychol.*, 1951, **42**, 257-276.

WALSH, ANN M. *Self-concepts of bright boys with learning difficulties.* New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1956.

(Received July 21, 1961)

Journal of Educational Psychology
1962, Vol. 53, No. 1, 43-47

PERSONAL PREFERENCE FOR METHOD AS A FACTOR IN LEARNING[1]

NEWTON E. JAMES
Student Counseling Bureau, University of Arizona

Do individuals who learn material utilizing a preferred method actually learn better than those utilizing a method other than their choice? A total of 503 basic airmen at Lackland Air Force Base were grouped according to expressed preference between reading and lecture as modes of receiving elementary Air Force technical training prose. Each group was then subdivided randomly, one segment being given learning material utilizing the mode of choice with the other segment utilizing a different mode. No differences were found on posttests between Ss receiving material in mode of choice and those receiving material in mode unlike that chosen.

The search for efficient ways in which to teach people to acquire new knowledge frequently leads over well-worn pathways, few of which have had more traffic than those supporting the investigation of modes of presentation of learning materials. For almost three-quarters of a century, there has been accumulating an extensive literature on the sensory and cognitive processes with much attention directed to the problem of accurately describing or comparing the learning which takes place when the learning material is presented visually and that which takes place when the material is presented in auditory form. (Brown, 1948; Carlson & Carr, 1938; Day & Beach, 1950; Elliott, 1936; Greene, 1928, 1934; Hollingworth, 1935; Stenius, 1945; Webb & Wallon 1956)

Since the early investigations of this type the technical advances in communications media have produced far reaching changes in the reading and listening habits of the American public. With the emergence of radio and television as mass media there were extensive changes in the amount of practice or informal training which persons received in reading and listening. The facilitation of learning which might be expected due to increased practice may have been offset by the cultivated inattention which people have acquired as members of the mass media captive audience (Meyersohn, 1957). Consequently, it is difficult to estimate what changes technological developments in communications have made in the ability of individuals to learn by reading or by listening.

A theoretical concept which has attained considerable popularity during the past decade, and the concept which prompts this study, contains the general idea that the image which a person has of himself is a motivating factor in his behavior and determines, to a certain extent at least, the conditions and extent of his learning (Benjamins, 1950; Lecky, 1951). By implication it follows that

[1] This article is based on a PhD dissertation submitted in 1958 to the Graduate School of the University of Minnesota. The author wishes to express his appreciation for the guidance received from his advisors, Ruth Eckert, Gordon M. A. Mork, and C. Gilbert Wrenn. The study program of which the dissertation is a part was sponosored by the Air Force Institute of Technology and the experimentation was conducted by personnel of the Air Force Personnel and Training Research Center. Acknowledgment of valuable assistance is made from these organizations and from Edwin S. Ewart, Marion E. Hook, and Dee W. Norton, all formerly of the latter organization.

when a person is given learning material in a manner which he thinks is advantageous to him, he will perform better than he would have performed had the material been presented in another manner. The specific problem to which this study is addressed is contained in the following question: Do individuals who are given learning material by the mode of their preference actually learn better than those individuals who are given the material in a mode unlike their choice?

PROCEDURE

Subjects were basic airmen recently inducted and in training at Lackland Air Force Base, Texas, in June 1957. The average age of individuals in the sample was 18 years and on the Armed Forces Qualifying Test (AFQT), a nationally standardized measure of general learning ability, the mean of the group was at the sixty-fourth percentile. The educational level of the group was not tabulated but in view of the high mean on the ability measure and the time of year at which the sample was drawn, it is inferred that the group contained many new high school graduates and also a number of college students who had chosen to interrupt their formal education at that time.

On each of 4 successive days subjects were received in groups for testing. These groups varied in size from 75 to 170 airmen; no individual appeared more than once. Each of the groups was treated in an identical manner and all testing was done between 7:30 and 9:00 A.M. in large, well-ventilated, well-lighted classrooms.

Each group was assembled in a large classroom where, after a brief introduction, a pretest of aviation information was administered. The purpose of this test was to determine the amount of prior knowledge which each individual had concerning the learning material which was to be presented in the experimental situation.

Following the pretest and a brief explanation of the study, the subjects were asked to indicate which of two modes of presenting technical learning material they preferred. The choices offered were: (*a*) I prefer a reading presentation, (*b*) I prefer a lecture presentation, and (*c*) I have no preference. Subjects were assigned then to groups according to the choices which they expressed. These groups are subsequently referred to as the choice groups.

Each choice group was then divided randomly into two groups of equal size and after additional introductory and explanatory remarks one group was given a written presentation of the learning material and the other group was given the same material over a public address system from a tape recording. Immediately following each presentation all groups were tested with a written multiple-choice test. These groups are subsequently referred to as the choice-treatment groups.

Scores on the AFQT, the general aptitude measure previously mentioned, were available for all the subjects. On the basis of these scores, the choice-treatment groups were each divided into two groups consisting of those individuals with scores above the group mean and those with scores below the mean. The division of the experimental group into subgroups is indicated in Figure 1.

The learning material and posttest were taken directly from a reading comprehension test which had been developed in the aviation psychology program of the Air Force. The test consisted of seven paragraphs of narrative prose, 200-400 words in length, on diverse technical scientific and historical subjects within the field of military aviation. Each paragraph was followed by a five-item multiple-choice question. The total num-

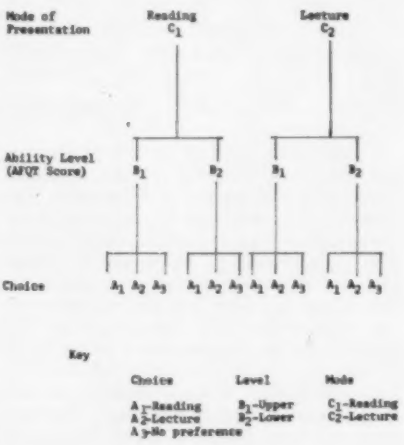


FIG. 1. Arrangement of groups for analysis.

ber of questions was 30 and the time allotted to the test when given in its original form as a silent reading test was 35 minutes.

The readability of the material, as measured by current formulae, was more difficult than that which persons of the age and educational level of basic airmen can be expected to read with 75% comprehension. However, the paragraphs of the material are typical of those found in textbooks for beginning technical courses in which basic airmen are enrolled.

The test was given to half of the subjects as a silent reading test and to the remainder of the subjects as a tape recorded presentation over a public address system. In each case the material was presented in revised form, one paragraph at a time, with the requirement that the subject read or listen to the material and immediately thereafter answer the questions pertaining thereto. All subjects were allowed 2½ minutes to read or listen to the material and 2½ minutes to answer the questions on each paragraph. Tests were scored on the basis of number of questions answered correctly. There was no correction made for guessing.

A three-dimensional factorial design was employed, and analysis of variance applied to both the pretest and posttest scores. The variables in the analysis were as follows:

1. subject's choice of mode of presentation of learning material, viz., *reading, lectures,* or *no preferences*
2. subject's AFQT score, viz., *above group mean,* or *below group mean*
3. actual mode of presentation of learning material, viz., *reading,* or *lecture* (tape recorded)

Null hypotheses were tested for each of the above three variables and for each possible interaction.

RESULTS

Analysis of the pretest data revealed no statistically significant differences between the experimental groups except those due to differences in AFQT scores. A summary of analysis of posttest data is given in Table 1.

Detailed description of the findings appears in the following paragraphs. Letters in parentheses refer to sources of variation reported in Table 1.

1. No differences were found in the scores of those subjects who preferred reading, those who preferred lecture, and those who stated no preference. (A)

2. Differences in scores indicated the superiority of individuals in the upper category of AFQT scores over those in the lower AFQT category. These differences are significant at the .01 level. (B)

3. For the sample as a whole, the reading presentation produced higher scores than did the lecture presentation. The differences are significant at the .05 level. (C)

4. There is no significant interaction between choice of presentation and category of AFQT score. That is to say, preference for a given mode does not appear to be any more potent factor for one level of ability than for the other. In neither group are these meaningful differences in scores between those who chose one method and those who chose another. (AB)

5. There is an interaction between the subjects' expressed choice and the mode of presenting the learning material which is statistically significant at

TABLE 1

POSTTEST DATA: SUMMARY OF ANALYSIS OF VARIANCE

Source of variation	*df*	*SS*	*MS*	*F*
A (Choice)	2	56.78	28.39	1.59
B (Level)	1	3181.31	3181.31	178.42*
C (Mode)	1	550.27	550.27	30.86*
AB	2	55.71	27.86	1.56
BC	1	67.84	67.84	3.80*
AC	2	115.94	57.97	3.25*
ABC	2	5.07	2.54	.14
Within cells	356	6346.98	17.83	
Total	367	10379.90		

*Significant at the .05 level (the maximum probability of Type I error).

the .05 level. This interaction does not operate, however, to produce significant differences in scores on the criterion measure between groups of subjects who received the learning material in the preferred mode of presentation and those who were given the material in a different mode or between these groups and the students who expressed *no preference*. (BC)

Means and standard deviations of the experimental groups appear in Table 2.

In each of these three groups, 42 subjects received the *written* presentation and 72 the *lecture;* half the subjects in each group were above the sixty-fifth percentile on national norms of the aptitude test employed and half were at or below this figure. There are no differences between any two of the above groups which are statistically significant at the .05 level.

6. The superiority of high ability over lower ability groups was more pronounced with the reading than the lecture presentation. This interaction between the aptitude test score and mode of presentation is significant at the .05 level of confidence. (AC)

7. There is no significant triple interaction between the variables. That is to say, there is no meaningful relationship between preference of mode of presentation, aptitude test score, and response to the mode of presentation employed. (ABC)

TABLE 2

SUMMARY OF SCORES

Group	Mode of presentation	*MS*	*SD*
1	In accordance with subject's choice	17.34	5.31
2	Contrary to subject's choice	17.10	4.96
3	Neutral with respect to subject's choice, i.e., subject stated *no preference*	18.00	5.54

CONCLUSION

1. When individuals are of equal intelligence, the learning of those who prefer the reading presentation is not significantly different from that of those who prefer the lecture presentation. However, as has been noted by other investigators, there is probably a difference in the achievement of people in general who choose the reading method and those who choose the lecture method. In the present study, it is noteworthy that apparently 40% of the total sample preferred the lecture presentation, 20% preferred the reading presentation, and 40% had no preference. Greene (1928, 1934), and other investigators have reported that the brighter students and the good readers preferred the reading method whereas the mediocre or duller students preferred the lecture. This finding seems to be supported in the present investigation, for 47% of the airmen who chose the reading presentation were in the upper category of the aptitude measure while only 38% of those who chose the lecture presentation were in the upper category. However, 52% of the airmen who expressed no preference were in the upper category of ability.

2. For the type of material which must be taught in Air Force technical schools the reading presentation has certain advantages that cannot be ignored. A certain amount of such material probably must be presented in this manner. The reading is difficult, as measured by formulae of reading ease, and involves unrelated factual items that require visual reference and convenient referability. The superiority of the reading presentation over the simu-

lated lecture presentation for the materials and subjects of this study is clearly established. It is likely that the substitution of a good, interesting lecturer for the tape recording would add considerably to the effectiveness of the lecture presentation. Yet, it seems equally likely that the substitution of a good, well-illustrated text or manual, complete with diagrams, figures, and index for the present written material would also increase the effectiveness of the reading presentation.

3. There appear to be no meaningful differences between individuals in the higher category of AFQT scores and those in the lower category with respect to the relationship between their choice of mode and their achievement. Choice of mode seems to make no difference in the criterion test scores in either AFQT category.

4. The findings suggest that when material is to be used for test purposes, a reading presentation brings out the differences between groups to a greater extent than does the lecture presentation.

5. Although the interaction between choice and mode of presentation is statistically significant at the .05 level, there are no significant differences between any two of the subgroups. The learning of the groups that received the material in the preferred mode was not significantly different from that of groups that received the material in a mode unlike that chosen. The differences between these definite choice groups and the groups that expressed no preference were likewise significant neither from the practical standpoint nor statistically at the .05 level. Consequently it is concluded that in comparison with other factors which affect the learning situation, whether or not a person gets his favored mode appears to be of relatively little importance. The variation in scores on the criterion measure which can be attributed to differences in general intelligence is over fifty times as great as that which can be attributed to the choice and mode interaction factor.

REFERENCES

BENJAMINS, J. Changes in performance in relation to influence upon self-conceptualization. *J. abnorm. soc. Psychol.*, 1950, **45**, 473-480.

BROWN, J. G. A comparison of listening and reading ability. *Coll. English*, 1948, **10**(2), 105-107.

CARLSON, H. B., & CARR, H. A. Visual and vocal recognition memory. *J. exp. Psychol.*, 1938, **23**, 523-530.

DAY, W. F., & BEACH, B. A survey of the research literature comparing the visual and auditory presentation of information. *USAF tech. Rep.*, 1950, No. 5921.

ELLIOTT, F. R. Memory for visual, auditory and visual-auditory material. *Arch. Psychol. NY*, 1936, No. 199.

GREENE, E. B. The relative effectiveness of lecture and individual reading as methods of college teaching. *Genet. psychol. Monogr.*, 1928, **4**, 459-563.

GREENE, E. B. Certain aspects of lecture, reading, and guided reading. *Sch. Soc.*, 1934, **39**, 619-624.

HOLLINGWORTH, H. L. *Psychology of the audience*. New York: American Book, 1935.

LECKY, P. *Self consistency: A theory of personality*. New York: Island Press, 1951.

MEYERSOHN, R. B. Social research in television. In B. Rosenberg & D. M. White (Eds.), *Mass culture*. Glencoe, Ill.: Free Press, 1957. Pp. 345-357.

STENIUS, A. C. Auditory and visual education. *Rev. educ. Res.*, 1945, **15**, 243-255.

WEBB, W. B., & WALLON, E. J. Comprehension by reading versus hearing. *J. appl. Psychol.*, 1956, **40**, 237-240.

(Received July 28, 1961)

Journal of Educational Psychology
1962, Vol. 53, No. 1, 48-51

VALUES OF MEDICAL SCHOOL STUDENTS AT DIFFERENT LEVELS OF TRAINING[1]

LEONARD V. GORDON
United States Naval Personnel Research Activity, San Diego, California

AND IVAN N. MENSH[2]
School of Medicine, University of California, Los Angeles

A cross-sectional study of changing values of medical school students was performed by administering a values test, the Survey of Interpersonal Values, to all members of the 1st through 4th year classes in a large medical school. A predicted decrease in Benevolence, from the 1st through the 4th year, was found. Significant increases in Support, Recognition, and Independence and a significant decrease in Conformity were also noted. The present findings have implications for the practice of applying personality data obtained on occupational groups to high school or college counselees.

A number of studies support the point of view, popularly held among medical school teachers, that the medical student, as he moves from his first through his fourth year, becomes less "idealistic" and more "cynical." In one of these, Eron (1955) surveyed attitudes of medical students and found that, year by year, students did, in fact, score higher on "cynicism" and lower on "humanitarianism." Becker and Geer (1958), in a 4-year longitudinal study of a group of medical students, using interview and observational techniques, also noted a shift away from idealism. They interpreted their findings as being indicative of the students becoming increasingly concerned with the technical aspects of their cases, and less interested in "helping the sick" or "sympathizing" with their patients. Kelley (Gee & Cowles, 1957) also noted this general lack of interest on the part of medical students in the welfare of human beings.

The Survey of Interpersonal Values (SIV) (Gordon, 1960) is a brief forced-choice test designed to measure the relative importance which one ascribes to each of six factored interpersonal value dimensions. One of the values included in this test, Benevolence, provides a measure which appears to be related to the humanitarianism or idealism referred to by Eron, and Becker and Geer. The six values are defined as follows:

Support—Being treated with understanding, receiving encouragement from other people, being treated with kindness and consideration.

Conformity—Doing what is socially correct, following regulations closely, doing what is accepted and proper, being a conformist.

Recognition—Being looked up to and admired, being considered important, attracting favorable notice, achieving recognition.

Independence—Having the right to do whatever one wants to do, being free to make one's own decisions, being able to do things in one's own way.

Benevolence—Doing things for other people, sharing with others, helping the unfortunate, being generous.

Leadership—Being in charge of other people, having authority over others,

[1] Opinions expressed are those of the authors and do not necessarily reflect those of the Department of the Navy.

[2] Grateful acknowledgment is made to Faye E. Barnes, Marielle Fuller, Ada Hirschman, Evelyn Kaslow, and Mary Linn Thomason for their assistance in the processing of the data.

being in a position of leadership or power.

The present study was performed to determine whether a downward trend would, in fact be found from the first through the fourth year on that aspect of idealism measured by the Benevolence scale of the Survey of Interpersonal Values. In addition, it was desired to determine whether trends would be observed for interpersonal values other than Benevolence, since it appeared to be reasonable to assume that other values also might be affected.

PROCEDURES AND RESULTS

The SIV was administered to all students in the first through the fourth year classes of a large western medical school. Two hundred and eight male and 14 female students were included in the testing. Freshmen were tested in the beginning of their first year; the other classes were tested during the course of their academic year. In view of the small size of the female sample, analysis was restricted to the male student group. Means and standard deviations for the six interpersonal values are presented, by class, in Table 1.

It was desired, first, to determine whether a downward trend existed from the first through the fourth year on the Benevolence value. Since it was reasonable to assume that any changes in values that did occur would be monotonic, following Lindquist (1953, p. 342), a significance test was performed between Benevolence means of the first and fourth year groups only. The mean difference between these two groups was found to be significant beyond the .01 level of confidence (Table 2), supporting the hypothesis that a downward trend occurs from year to year on the Benevolence scale.

Tests of the significance of the differences between means of the two classes, for the other five values, were performed next. Mean differences were found to be statistically significant (Table 2) for four out of

TABLE 2

MEAN DIFFERENCES AND SIGNIFICANCE LEVELS

Value	Fourth year—First year		Resident—First year	
	Mean difference	t	Mean difference	t[a]
S	3.03	2.88**	1.29	.62
C	−3.28	3.06**	−5.64	2.91**
R	2.79	2.01*	2.45	1.66*
I	3.54	2.73**	4.30	1.70*
B	−5.01	5.12*	−5.86	3.55**
L	−1.19	.92	3.63[b]	1.85

[a] One-tailed test.

[b] The difference for Leadership is in the unexpected direction.

* Significant at .05 level.

** Significant at .01 level.

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR FIRST THROUGH FOURTH YEAR MALE MEDICAL STUDENTS AND MALE PSYCHIATRIC RESIDENTS ON THE SIV

Group	N	Means						Standard deviations					
		S	C	R	I	B	L	S	C	R	I	B	L
First year	66	13.39	12.97	10.88	18.26	18.08	16.26	5.69	6.34	4.83	6.77	4.98	6.14
Second year	49	17.86	9.84	12.27	20.57	15.57	13.84	4.35	5.37	3.60	7.17	5.00	4.55
Third year	48	17.31	8.21	11.77	21.77	14.67	16.44	4.73	5.05	4.48	5.75	4.99	4.95
Fourth year	45	16.42	9.69	13.67	21.80	13.07	15.07	5.19	4.83	4.68	6.54	5.04	6.95
Residents	10	14.68	7.33	13.33	22.56	12.22	19.89	5.86	5.31	4.06	7.16	4.60	5.44

Note.—S = Support, C = Conformity, R = Recognition, I = Independence, B = Benevolence, L = Leadership.

the five remaining values. Fourth year students place a higher value on Independence, Recognition, and Support than do first year students, but value Conformity less. No significant difference was found between classes on Leadership.

A group of 10 psychiatric residents, who had just completed their internship, and who were at the same medical facility as the medical students, were also administered the SIV. Means and standard deviations for this group are presented in Table 1. As a matter of interest, significance tests were run to determine whether differences previously found between the first and fourth year groups would also be found between the first year and resident groups. Since a particular directionality was being tested for each difference, one-tailed tests of significance were used. As may be noted in Table 2, for all five values on which significant mean differences had been previously found, differences between first year students and residents were in the same direction as the previous differences. Four of these differences were significant at or beyond the .05 level of confidence.

It may be noted that sharper differences in mean value scores occur between first and second year students than between other consecutive year groups. These differences are not due to first year attrition, since, for the present freshman class only four students failed and differences between classes with these students excluded are essentially the same as those presently found.[3] The larger differences between first and second year students are in accord with Freedman's findings (1961), that personality changes during college training are not linear from year to year but tend to take place quite early.

DISCUSSION

The present study confirms the findings of other investigators that from the first year on, being benevolent, in the sense of wanting to help other people, becomes decreasingly important to the medical student. This downward shift in Benevolence is consistent throughout medical school training and seems to carry on beyond graduation. Whether this trend continues into medical practice, of course, cannot be ascertained from the present study.

The decline in Benevolence exhibited by medical students during their course of training need not be too disconcerting. The average entering medical student has a percentile rank on Benevolence of 70 on male college norms, 75 on engineering norms, and 84 on middle management norms. In his fourth year of medical training, his percentile rank is now 44, 47, and 54, respectively, on these three sets of norms. Thus while the beginning student is, indeed, rather idealistic in his desire to help his fellowman, as graduation nears, he is probably no less benevolent than the average adult male.

Equally interesting are other systematic differences which are suggested by the present data. The most consistent trend, other than the decrease in Benevolence, is found in the medical student's increased valuing of Independence as he progresses through medical school training. In addition, an increase in his valuing of Recognition may be noted. On the other hand, conforming behavior becomes less important to him. The drop in both Benevolence and Conformity may reflect an increase in problem orientation, and a decrease in social orientation, on the part of the medical student, as Becker and Geer have suggested. He may feel the need for a more impersonal relationship toward his patients, as well as a greater freedom from societal restraints. While these are a posteriori speculations, they do present hypotheses that merit further investigation if a broader understanding of value changes which occur during medical school training is to be achieved.

Present findings should also caution the

[3] Differences between means of all students and passing students averaged .1. The sample size of the first year group reflects a 50% expansion in the freshmen class over previous years.

guidance counselor against the uncritical use of personality profiles obtained on job incumbents with high school or college freshman counselees. Successful or vocationally adjusted job incumbents may have quite different traits or values from what they had prior to their academic or vocational training. Utilization of personality profile data for vocational guidance purposes should follow a prior determination of the stability of relevant dimensions by cross-sectional or, preferably, longitudinal investigations.

REFERENCES

BECKER, H. S., & GEER, BLANCHE. The fate of idealism in medical school. In E. G. Jaco (Ed.), *Patients, physicians and illness.* Glencoe, Ill.: Free Press, 1958. Ch. 32.

ERON, L. D. Effect of medical education on medical students. *J. med. Educ.*, 1955, **10**, 559-566.

FREEDMAN, M. B. Influence of college experience on personality development. *Psychol. Rep.*, 1961, **8**, 21-22.

GEE, HELEN H., & COWLES, J. T. (Eds.) *The appraisal of applicants for medical school.* Evanston, Ill.: Association of American Medical Colleges, 1957.

GORDON, L. V. *Survey of Interpersonal Values.* Chicago: Science Research Associates, 1960.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

(Received September 15, 1961)

a child

is this.....and this

3 NEW TEXTS ON HOW HE DEVELOPS, HOW HE LEARNS

CHILD DEVELOPMENT AND ADJUSTMENT:

A Study of Child Psychology

By Lester D. Crow, Brooklyn College, and Alice Crow

This text covers every significant aspect in the development of the normal child, from pre-natal stages through early adolescence, incorporating the results of the latest research studies. Each aspect of development is treated as an integral part of the development of the whole child. Physical, mental, and emotional growth, as well as motor, language, and social development are discussed in relation to each other and to the total personality pattern. In addition, materials are continually correlated with other areas of psychology.

1961, 515 pages, $6.50

EDUCATIONAL PSYCHOLOGY,

Revised Edition

By Glenn Myers Blair, R. Stewart Jones, and Ray H. Simpson, all, University of Illinois

The latest edition of this text covers the facts, principles, and methods that lead to a sound psychological foundation for effective teaching. It features up-to-date research on teaching machines, feedback, "brainstorming," creative thinking, and discovery and learning. Pictures of classroom learning activities, lists of new instructional films, and end-of-chapter exercises offer concrete suggesions, and a teacher's manual provides excellent testing devices.

February, 1962

READINGS IN EDUCATIONAL PSYCHOLOGY

Edited by Victor H. Noll and Rachel P. Noll, both, Michigan State University

A completely up-to-date book of readings designed to supplement any standard text, this volume includes selections on growth and development, learning, measurement and evaluation, mental hygiene, and the science of educational psychology. Alternate viewpoints are presented on major topics, and the emphasis is on articles that incorporate the latest research studies.

April, 1962

THE MACMILLAN COMPANY
60 Fifth Avenue, New York 11, New York

A Division of
The Crowell-Collier Publishing Company

Psychology and Mental Health

A report of the Institute on Education and Training for Psychological Contributions to Mental Health, held at Stanford University in August, 1955.

Edited by CHARLES R. STROTHER

Topics discussed at the Institute:

Training Needs of Psychologists in Community Mental Health

Specialization in Training

Practicum Training

Training for Therapy

Training for Research in the Mental Health Field

Problems of Departmental Organization

Price: $1.75

Order from:

American Psychological Association, Inc.
1333 Sixteenth Street, N. W.
Washington 6, D. C.

two

outstanding
books from
Wadsworth:

PSYCHOLOGY A Scientific Study of Man
by Fillmore Sanford, University of Texas

This brilliant textbook for the beginning student of psychology presents with clarity and excitement challenging material that imparts an intellectual appreciation for and a detailed understanding of psychology. Emphasis is placed on the *methods* of psychology and their relation to the intellectual and social life of our culture.

Clothbound 8 x 10½" 588 pages text list $7.50
Workbook Paperbound 8½ x 11" 256 pages text list $2.50
Teacher's manual available upon adoption.

EDUCATIONAL PSYCHOLOGY
by Frederick J. McDonald, Stanford University

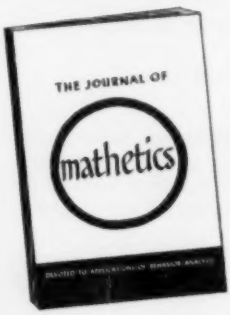
Within a very short time this book has become a leader in the field of educational psychology textbooks. Its new approach emphasizes the teacher as a hypothesis maker—as one who uses the framework of psychology to formulate, test, and evaluate teaching methods related to the development of personality in the child.

Clothbound 6¾ x 9¼" 768 pages text list $6.95
Teacher's manual available upon adoption

For further information you are invited to write

Wadsworth Publishing Co.
Belmont, California

ANNOUNCING ***The Journal of Mathetics***



THOMAS F. GILBERT
Editor in Chief

CARROLL DUNN
Executive Editor

Editorial Advisors

JOHN A. BARLOW CHARLES B. FERSTER PAUL S. SIEGEL
JOSEPH V. BRADY ISRAEL GOLDIAMOND CHARLES W. SLACK

Papers of profound interest to educators on new applications of reinforcement theory are becoming available in growing numbers.

The Journal of Mathetics fills the need for an organ devoted to their collection.

It gives the scientists in this field a publication of professional standing in which they may discuss their own work and keep informed on the findings of their contemporaries.

Bi-monthly from January 1962

Annual subscription .....$18.
A.P.A. member rate .....$12.
A.S.T.D. member rate .....$12.
Student rate .....$10.

The Journal of Mathetics **Box 3232 E University, Alabama**

ELEMENTARY STATISTICAL METHODS IN PSYCHOLOGY AND EDUCATION

Paul Blommers and E. F. Lindquist

"This seems to me an excellent elementary text which maintains the balance I think necessary between mathematical vigor and intuitive argument. . . ."—Jacob Cohen, New York University. *528 pp. 1960 $6.00. Accompanying Study Manual and Instructor's Key.*

PSYCHOLOGY IN TEACHING AND LEARNING

William Clark Trow

". . . Mr. Trow has done what I consider an excellent job both from the standpoint of interesting material and design. . . ."—E. O. Wood, Baylor University. *488 pp. 1960 $5.75. Accompanying Student's Manual and Instructor's Handbook.*

Houghton Mifflin

BOSTON · NEW YORK · ATLANTA
GENEVA, ILL. · DALLAS · PALO ALTO

Wiley BOOKS

Just out — the new SECOND EDITION of

EDUCATIONAL PSYCHOLOGY IN THE CLASSROOM

By HENRY CLAY LINDGREN, *San Francisco State College.* A sound, thorough revision of a popular, *classroom-oriented* text; in this edition Lindgren sharpens his idea of the teacher functioning as a behavioral scientist. Emphasizing the social and emotional factors that affect learning, the book is firmly based on laboratory experience and research, and reflects the most modern thinking on the subject. *New in the Second Edition:* increased attention to motivation and the under-achiever . . . stress on secondary school problems . . . revision of testing and evaluation material . . . more coverage on special education and automated teaching . . . a new chapter on discipline. Student workbook and instructor's manual available.

1962 574 pages $6.95

Send now for your on-approval copy

JOHN WILEY & SONS, Inc.

440 PARK AVENUE SOUTH, NEW YORK 16, N. Y.